

Spring 5-15-2015

## Stories of Principals' Roles as a Result of High Stakes Teacher Evaluations in Louisiana

Jenna L. Galjour Chiasson  
*University of New Orleans*, [jgaljour@uno.edu](mailto:jgaljour@uno.edu)

Follow this and additional works at: <https://scholarworks.uno.edu/td>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

---

### Recommended Citation

Galjour Chiasson, Jenna L., "Stories of Principals' Roles as a Result of High Stakes Teacher Evaluations in Louisiana" (2015). *University of New Orleans Theses and Dissertations*. 1965.  
<https://scholarworks.uno.edu/td/1965>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Dissertation has been accepted for inclusion in University of New Orleans Theses and Dissertations by an authorized administrator of ScholarWorks@UNO. For more information, please contact [scholarworks@uno.edu](mailto:scholarworks@uno.edu).

Stories of Principals' Roles as a Result of High Stakes Teacher Evaluations in Louisiana

A Dissertation

Submitted to the Graduate Faculty of the  
University of New Orleans  
in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy  
in  
Educational Administration

by

JennaLynn Galjour Chiasson

B.A. Nicholls State University, 2009  
M.Ed. Nicholls State University, 2011

May, 2015

Copyright 2015, JennaLynn G. Chiasson

## **Dedication/Acknowledgements**

As many times as I have dreamed of composing this page of my dissertation, there has always been a doubt that the day would ever come. This journey was begun and accomplished all because mentors along the way saw more in me than I ever saw in myself. Dr. Leslie Jones pushed me as an undergraduate student to continue my education; when I called UNO, Dr. D encouraged me to apply and helped me through the process; my wonderful chair Dr. Beabout stepped in and supported me through the writing process. I will always be thankful for meeting my dear friend and peer, Diana Ward, on my very first night of class at UNO. Through an engagement, a wedding, two babies, a full-time job, and a full-time course-load, no one at UNO ever told me I couldn't do it, and for that I will always be grateful.

My fantastically supportive family helped me through all of it to the point where I feel like I should tear my diploma and distribute it because I never could have done it without them. My parents have been the most amazing parents anyone could ever ask for. Their willingness to give of themselves to their children inspires my brothers and I to strive for seemingly unreachable goals in our efforts to make them proud. Thank you to my in-laws, my brothers, and everyone else in my large and loving family who were always there for me whenever I needed them.

To my husband, Grant: you have saved me from myself every day since I have met you. Every time I think I cannot do something, you somehow always know that I can. My world is a beautiful place because of what you have given me.

To my babies, Gavrie Paul and Baby 2: you have literally accompanied me on this journey and inspired me in ways you will probably never understand. I dedicate this dissertation to you both because I did all of it for you.

## Table of Contents

List of Tables .....	vii
Abstract.....	viii
Chapter 1: Introduction .....	1
Personal Reflection: My Own Introduction to COMPASS.....	2
Problem Statement .....	2
Purpose.....	3
My Subjective I's Revealed .....	4
A Brief History of the Literature on Teacher Evaluation .....	5
Research Question.....	9
Chapter 2: Policy Context and Literature Review .....	11
Introduction.....	11
Context of the Study.....	12
State Policy.....	12
The School District.....	15
Review of Literature .....	16
History of Teacher Evaluation.....	16
Value-Added Measurement: Quantitative Teacher Evaluation.....	20
Standards-Based Teacher Evaluation Systems: Qualitative Teacher Evaluation.....	24
The Principal's Role in Teacher Evaluation: Problems and Issues .....	27
The Principal's Role in Teacher Evaluation: Principals' Perceptions .....	30
The Principal's Role in Teacher Evaluation: Instructional Leaders.....	31
Theoretical Framework.....	33
An Open System .....	33
Contingency Theory .....	34
Instructional Leadership .....	36
Why These Theories.....	38
Chapter 3: Methods .....	40
Personal Reflection: My I as an Evaluator.....	40
Introduction.....	41
Purpose.....	41
Research Question.....	41

Research Design .....	42
Participants .....	43
Data Collection.....	44
Gaining Access .....	44
Sampling .....	44
Data Collection.....	45
Data Analysis .....	45
Narrative Analysis.....	45
Thematic Analysis .....	47
Verification.....	47
Member Checking.....	47
Peer Debriefing.....	48
Ethical Considerations.....	48
The Researcher’s Role.....	49
Field Relations .....	49
Reflexivity.....	49
Reflective Writing.....	49
My Story .....	50
Chapter 4: Findings .....	52
Table 1: Profile of Participants.....	53
Narrative Analysis: Stories of Principals as Evaluators.....	54
Mr. Boudreaux’s Story .....	54
Mr. Breaux’s Story .....	54
Ms. Robichaux’s Story.....	63
Mr. Richard’s Story.....	66
Ms. Simoneaux’s Story .....	69
Mr. Thibodeaux’s Story .....	71
Ms. Fusilier’s Story .....	73
Thematic Analysis: The Roles of Principals Implementing COMPASS.....	78
Instructional Coach.....	79
High Stakes Evaluator.....	85
Revisiting my Subjective I’s .....	91
Summary .....	92
Chapter 5: Discussion.....	93

Revisiting the Roles.....	93
Instructional Coach.....	93
High-Stakes Evaluator .....	100
Implications for Theory.....	105
Implications for Practice.....	107
Implications for Policy.....	109
Limitations of the Study and Future Research.....	111
References .....	113
Appendix A .....	119
Appendix B .....	120
Appendix C.....	124
Appendix D .....	125
Appendix E.....	126
Appendix F.....	127
Appendix G .....	128
Vita.....	129

## **List of Tables**

Table 1: Profile of Participants.....	57
Table 2: Use of Walkthrough Observations.....	85



## Abstract

In 2010, the Louisiana legislature passed Act 54, a law that requires public school teachers to undergo a performance-based accountability evaluation. COMPASS (Clear, Overall Measure of Performance to Analyze and Support Success) asks principals to evaluate teachers using a rubric with components of Charlotte Danielson's *Framework for Teaching* to evaluate teacher effectiveness (Act 54). Act 1, passed in 2012, ties Louisiana public school teacher's pay and tenure to their score on COMPASS. Principals of Louisiana are now asked to evaluate teachers in a high stakes evaluation that is linked to teacher tenure and pay.

A qualitative study using narrative research design was conducted to explore how principals described their roles as high-stakes evaluators through the implementation of COMPASS. Data was collected from seven participants in the form of in-depth interviews and each was recorded and transcribed for data analysis. Restorying and story mapping were used to compose narratives that describe the roles of the participants in their implementation of COMPASS. Using the theoretical frameworks of Contingency theory and Instructional Leadership theory, two roles emerged from their narratives: Instructional Coach and High Stakes Evaluator. The information gleaned from this study can help to inform future policy about possible issues with COMPASS in implementation as well as impact future practice for evaluators from the stories of the participants.

KEYWORDS: teacher evaluation, principal, education policy, Louisiana, leadership

## **Chapter 1: Introduction**

### **Personal Reflection: My Own Introduction to COMPASS**

I first began the journey of studying the COMPASS teacher evaluation system in the spring of 2012 as I took a policy analysis course. At the time, Louisiana legislation was passing a wave of new education reform policies, including policies about teacher tenure and pay for performance based on value-added scores and evaluation using the new COMPASS rubric. Fast forward to the fall of 2012 when I was sitting in a professional learning community meeting at my school in which we were discussing the new COMPASS rubric we were to be evaluated on. As a fourth year teacher who was very motivated to move my students, I had always pushed myself to learn as much as I could. I understood the majority of the rubric, but I distinctly remember asking my head principal and one assistant about a part of the rubric that I did not understand. I did not understand a piece under the “Assessment” category: “Students appear to be aware of, and there is some evidence that they have contributed to, the assessment criteria. Students self-assess and monitor their progress” (LDOE, COMPASS rubric). I asked the two principals in the room what this would look like in a classroom because I didn’t know how to have students contribute to assessment criteria and I did not know how to have students monitor their own progress other than recording their grades. I was told that I shouldn’t be looking in the “Highly Effective” category because it was impossible, unattainable, and unrealistic. I should strive to just be proficient. That was my answer. It was at this point that I realized something was wrong. I, as a teacher, wanted to do my best for my students but I was not being given what I needed because it was obvious my principals had not been given what

they needed. This study is not an attack on COMPASS or on teacher evaluation. It is not research to prove value-added is an inadequate measure of teacher effectiveness. It does not seek to place blame on policy-makers, state leaders, or school principals. It simply strives to tell the stories of principals as they use the COMPASS evaluation system. If we want teachers to do better, we need to examine school leaders' use of the evaluation process and the roles they play within it. It is through this that we can inform future policy decisions and improve the practice of evaluators.

### **Problem Statement**

Interest in teacher evaluation has grown since the implementation of *Race to the Top*, which calls for "rigorous, transparent, and fair evaluation systems...that take into account data on student growth" (U.S. Department of Education, 2009, p. 9). Because of this, many states have adopted value-added systems containing both data on student growth and evaluation rubrics. States competed for federal money, and although Louisiana did not qualify, the state still adopted state policy for teacher evaluation. In the Spring of 2010, the Louisiana legislature passed Act 54, a law that requires public school teachers to undergo a performance-based accountability evaluation which includes a score made up of student growth data and classroom observations (Act 54). The COMPASS (Clear, Overall Measure of Performance to Analyze and Support Success) evaluation system was designed by the Louisiana Department of Education and was piloted briefly in the Spring of 2012 in nine districts. This system was based on Charlotte Danielson's research-based Framework for Teaching. Before full implementation in the fall of 2012, the Louisiana Department of Education switched to an abridged version of Danielson's Framework for Teaching. Charlotte Danielson helped Louisiana to create the COMPASS rubric which uses some of the

components of her *Framework for Teaching*, but she advises that “it decreases accuracy” and recommends using “the full instrument”; she was also unaware that the model would be used immediately without “working out the bugs” (Garland, 2012).

Recent teacher evaluation policy has set a precedent for a new type of school leader. School leaders are now required to evaluate classroom instructional practices in high-stakes teacher evaluations. Because of this, school leaders are asked now, more than ever, to become instructional leaders. Although instructional leadership models have existed since the 1980s (Hallinger & Murphy, 1985), components of school leaders in the role of evaluator are absent from them. Now that conducting teacher evaluations is a major part of a principal’s responsibilities, it is crucial to understand how principals fulfill the role of evaluator. In many modern teacher evaluation systems, school leaders must be able to objectively observe many grade levels and content areas, evaluate the effectiveness of the teacher, and provide relevant and supportive feedback to the teacher, thus demanding a school leader possess traits of an instructional leader. By exploring the stories of school leaders, we can learn more about how they are adapting to the policy changes in teacher evaluation. These stories can also inform implications for large-scale policy changes and decisions for implementing teacher evaluations at the district and school levels.

### **Purpose**

Charlotte Danielson works with state departments of education and school districts to train teacher evaluators to use her *Framework for Teaching*, a popular teacher evaluation model, and she lists one challenge as a need for trained evaluators:

A credible system of teacher evaluation requires higher levels of proficiency of evaluators than the old check-list, “drive-by” observation model. Evaluators need to

be able to assess accurately, provide meaningful feedback, and engage teachers in productive conversations about practice (Danielson, 2011, p. 38).

This is a call for instructional leaders because in order to fulfill these roles, school leaders must be well-versed in teaching and learning as a process. Louisiana's school leaders are being asked to fill the role of instructional leader in a way that many have never been asked to before. The narrative study proposed here will examine how principals have addressed this changing set of role expectations.

Blase and Blase (1999) examined teachers' perspectives of school leaders' instructional leadership characteristics and also calls for further research on the characteristics of school instructional leaders:

"The findings of our study as well as the emergence of diverse related issues in the literature suggest the fruitfulness of further study...study of principals' personal characteristics as well as political factors that may influence instructional leadership orientations" (Blase & Blase, 1999, p. 139).

As Blase and Blase (1999) suggest, a study of "personal characteristics" needs to be conducted to examine how these influence instructional leadership. This narrative study uses stories to garner school leaders' integration of personal abilities, role conception, and understanding of the policy. This will impact the practice of school leaders as well as give further implications for the policy development of teacher evaluation.

### **My Subjective I's Revealed**

The use of school leaders' stories of implementing COMPASS through a narrative methodology will allow for "retrospective meaning making" (Chase, 2005, p. 656). The stories will be interpreted and analyzed for thought processes of principals as they

navigate their complex roles, including that of evaluator. Because I am an educator, currently an employee of a public school district in Louisiana as a central office employee, subjectivity was an unavoidable factor in the collection, analysis, and interpretation of data throughout this study. For this reason, the use of an “interactive voice” was employed by the researcher in order to understand myself and my narrators (Chase, 2005, p. 666). I must tell my stories and personal reflections to reveal my own subjectivity, which “is an invariable component of [my] research” (Peshkin, 1988, p. 17). Peshkin argues for qualitative researchers to “uncover” their “subjective I’s”:

Researchers should systematically seek out their subjectivity, not retrospectively when the data have been collected and the analysis is complete, but while their research is actively in progress. The purpose of doing so is to enable researchers to be aware of how their subjectivity may be shaping their inquiry and its outcomes (Peshkin, 1988, p. 17).

Within further chapters of this study, I will reveal the I’s that I have uncovered in the process of designing this research, and I will include stories from my own personal experiences with COMPASS to help shape how these I’s opinions were shaped by those experiences.

### **A Brief History of the Literature on Teacher Evaluation**

Teacher evaluation is not a new area of interest to the field of education, although high-stakes summative evaluations are new to the state of Louisiana. Policies to evaluate teachers can be traced back to 1913 when Joseph Taylor first created rating scales to judge teachers’ influence on students, teaching ability, enthusiasm, discipline, and energy (Callahan, 1962). From here, teacher evaluation became increasingly thorough through the

decades as teacher quality has been identified as the single most important factor in determining student success (Odden, 2004; Nye, Konstantopoulos, & Hedges, 2004; Borman & Kimball, 2005; Kimball, et al., 2004; Odden, et al. 2004). Since *A Nation at Risk* (1983) called for district use of teacher evaluations for salary, promotion, tenure, and retention decisions the stakes of teacher evaluation have risen. Wise et al. (1985) studied teacher evaluation and posed several questions about the use of teacher evaluation that are still salient today, including the idea that principals should be relieved of some managerial tasks in order to have adequate time for more powerful observations and conferences, master teachers should be used to strengthen the evaluation process and give support to teachers, and teacher evaluation should be a district commitment that formulates specificities of the system to meet individual goals without state-imposed highly prescriptive teacher evaluation systems (Wise et al., 1985).

Originally, teacher evaluations were similar across the nation, a pass/fail checklist evaluation conducted once or twice a year by the school leader that the vast majority of teachers passed. *No Child Left Behind* (2001) brought to the forefront a focus on high-stakes standardized testing of students to rate student achievement and the “highly qualified” status of teachers which valued teacher knowledge of and experience in their content area(s). *Race to the Top* (2009) gave grant funds to states that applied for innovative projects to improve education. A stipulation of this grant funding was for states to use student achievement data within its teacher evaluation systems, referred to as value-added measures (United States Department of Education, 2009). Many states, including Louisiana, passed state laws to include this type of teacher evaluation systems and retained these policies even though they were not awarded the federal funding.

Value-added models measure the gain in a student's scores during a certain period of time. Alicia (2005) asserts that the value-added model "appears flawed essentially because it assumes that the gain score of students is attributable only to the teacher(s)" (p. 1). The selection of which student data to use has caused problems in the reformation of teacher evaluation (Ovando & Ramirez, 2007). In Louisiana, the shift to the Common Core State Standards in assessment has made it even more challenging to use student test data to formulate teachers' value-added scores. Value-added measurements in many states, including Louisiana, control for external factors but still draw much criticism from educators and testing experts. Stumbo and McWalters (2011) contest that standardized exams usually do not contain enough data to accurately determine a teacher's impact on student achievement. There is also argument about use of value-added measurements for the grades and subjects that do not have standardized tests (Stumbo and McWalters, 2011). Louisiana has adopted the use of teacher-constructed Student Learning Targets (discussed in more depth later) to address this issue.

The use of the data from teacher evaluations has changed over the years as well. Originally, teacher evaluations were used only as a way to give teachers feedback to improve their instruction and many were viewed as formative "checks" for quality teaching. Now, the idea of value-added measurements brings with it the concept of merit-based pay (Kimbal & Milanowski, 2009). In Louisiana, each individual district has been given the autonomy to devise a new salary schedule based on years of experience, demand, and level of effectiveness as defined by one's COMPASS score. This greatly differs from the old scales that are based on level of certification, years of experience, and level of education (LDOE, Act 1 Compensation).



Principals play a critical role in the teacher evaluation process; however, Jacob and Lefgren (2008) found that “one should not rely on principals for fine grained performance determinations as might be required under certain merit pay policies” (p. 129) because of multiple factors that could inhibit teachers from a pay increase unjustly. Rogers and Weems (2010) note that principal observation in the most common form of principal evaluation in schools, but they are limited in time and frequency causing difficulties for true teacher improvement; thus, their study argues for a more comprehensive approach to teacher observations to promote instructional growth.

Multiple problems with principals as teacher evaluators have emerged in recent years. Milanowski & Heneman (2001) identify three areas of concern: lack of subject matter knowledge, failure to provide helpful feedback, and inconsistency among evaluators. The National Education Association (2010) reports that the majority of principals have not been trained in evaluation methods to the degree at which they can provide feedback to teachers. Teachers will only use feedback from administrators when they believe their evaluator is skilled and competent. Donaldson (2009) discovered that administrators must commit to training because “without high-quality professional development, evaluators will not evaluate accurately and the evaluation will like have little impact on teaching or learning” (p. 11). Donaldson and Donaldson (2012) reported that a lack of trust and a lack of pedagogical knowledge inhibited the evaluation process from being a learning experience for teachers. In order for teachers to find their principal evaluations “credible and respond to them with efforts to build on their strengths and address their weaknesses, they must trust the observer and have access to subsequent learning opportunities” (Donaldson & Donaldson, 2012, p. 80). Inter-rater reliability also

comes into question as school leaders may all view the same evidence and still rate teachers differently. Danielson (2011) concurs that “even after training, most observers require multiple opportunities to practice...to calibrate their judgments with others” (p. 38). Teachers must believe their school leaders have sufficient training, a vast knowledge of both content and pedagogy, and have worked on inter-rater reliability with other evaluators in the school building and district in order to trust their evaluator when given feedback and opportunities for growth. Without these pieces, the evaluation process becomes nothing more than the assignment of a score.

Although the purpose of most modern teacher evaluation programs is to increase teacher effectiveness, many factors can encumber this objective. Today’s principal is asked by new teacher evaluation policy to shoulder the roles of evaluator and coach simultaneously through teacher evaluation. There is a need to discover, through principals’ stories, how principals interpret and implement teacher evaluation in their schools. The methods by which they are adapting to this new teacher evaluation system can be used to guide school leaders in navigating their role as teacher evaluator.

### **Research Question**

The following main research question is explored in this study: *How do principals describe their roles as high-stakes evaluator through the implementation of COMPASS?* Principals’ descriptions of their roles were discovered through their own stories told in interviews. The stories of these principals are central to this study through the use of a narrative methodology. Although teacher evaluation has been studied for over three decades, there has been an identified need to discover how principals describe their own roles as evaluators (Blase and Blase, 1999). Through these stories, principals relate how

COMPASS activities are integrated into their own approaches to leadership. One of the stated goals of COMPASS is to provide teachers with feedback to improve their instruction. Learning about the roles of school leaders as teacher evaluators will benefit practitioners currently utilizing COMPASS or other teacher evaluation programs, and it will inform policy-makers about the application of teacher evaluation policy at the school level.

## **Chapter 2: Policy Context and Literature Review**

### **Personal Reflection: My I as a Teacher**

I will always identify as a teacher. It is still my automatic reply when asked what I do, until I self-correct and explain I now work in the curriculum department of a central office for a public school district. With a mother, mother-in-law, and father-in-law who are all lifetime educators and the majority of my friends being teachers, it is difficult to extract myself from the role of a teacher. I'm not even sure that I should stop identifying as a teacher. With that comes my I as a teacher. I hear stories weekly about COMPASS and unfair administrators who don't know how to evaluate. Every semester, I calm my mom before her COMPASS evaluations, which send her into a panic for some reason even though she consistently is one of the top scoring teachers at her school. Throughout this study, I have had to remind myself of this I that lurks, seeking to blame administrators for inconsistencies and unjust evaluation results. I have had to step back into a more objective voice as a researcher and not seek to first blame the evaluator, or in this case, my participants.

### **Introduction**

This chapter includes an overview of the recent state policy on teacher evaluation in the state of Louisiana as well as information about the school district that the study took place in. Following the policy context and information about the school district is a review of literature of teacher evaluation that includes the history of teacher evaluation, value-added teacher evaluation, standards-based teacher evaluation systems, the principal's role in teacher evaluation, and conceptual framework.

## Context of the Study

### State Policy

Louisiana's educational system has historically been presented as substandard and in need of drastic reform. Louisiana's *Quality Counts* ratings the past three years have shown a critical need for progressive reform: although the state's national ranking rose from 35<sup>th</sup> in 2009 to 21<sup>st</sup> in 2011, Louisiana dropped to an F rating in K-12 student achievement in 2011 and has maintained that F rating in 2012 even though Louisiana's overall rating has risen, placing Louisiana in the 15<sup>th</sup> spot (*Quality Counts*). Republican Governor Bobby Jindal was reelected based in part on his education reform package including subsidies for school vouchers, increased accountability for school leaders and teachers, and school-based management. In an effort to pacify teacher unions and citizens, commercials were aired in the Spring of 2012 around the state claiming that "throwing money" at Louisiana schools is not the answer. Jindal proclaimed that "forty-four percent of our public schools are failing. 225, 000 are below grade level, and our state is spending a billion dollars a year on failing schools. That's unacceptable" (American Press, 2012). Jindal's efforts at educational reform began with the passage of Act 54.

In 2010 the Louisiana state legislature passed Act 54, a policy focusing on evaluating teachers and administrators based on student growth and classroom observations. According to Act 54, value-added scores serve as a quantitative measure encompassing fifty percent of teacher evaluations, and principal observations occupy the remaining fifty percent as a qualitative measure. Act 54 utilizes COMPASS (Clear, Overall Measure of Performance to Analyze and Support Success) as a model to evaluate educators. Educators

who are deemed ineffective based on these measures are to be placed on an intensive assistance program and then either be reinstated or face disciplinary action. For those teachers who fall into the Non-Tested Grades and Subjects (NTGS), state-approved common assessments were developed and serve as the quantitative portion of those teachers' evaluations. All teachers are required to write Student Learning Targets (SLTs) based on data from state-approved common assessments, and goal-attainment is the measurement used for those NTGS teachers. Because state standardized tests are now shifting to become Common Core State Standards aligned, teachers may not be receiving value-added scores for the 2013-2014 school year and perhaps several other years as the assessments shift. If this becomes the case, all teachers' quantitative scores will be derived from SLT attainment (Act 54, 2010).

Following Act 54, House Bill 974 was approved in April 2012 as a law that utilizes the data spawned from the Act 54 model to determine teacher tenure and pay for performance. House Bill 974, signed into effect as Act 1 included the following areas of change: employment contracts and personnel matters, salaries of teachers and other school employees, tenure, and termination of employment. Act 1 also included the right for teachers to challenge the qualitative measure, principal observations, with the quantitative portions of their evaluations, the value-added score. This huge bill includes evaluation for superintendents, administrators, and teachers, with each position to be measured by effectiveness over seniority or tenure (Act 1, 2012). According to this bill, teacher tenure earned before September 1, 2012 remained, but any untenured teachers following this date must be rated "highly effective" five out of six years in order to earn tenure. A tenured teacher who is rated "ineffective" shall immediately lose tenured status. "Effective" and

“ineffective” ratings are based on COMPASS (passed in Act 54, 2010). Teacher salary schedules had to be updated by each individual district by January 1, 2013 and had to be in effect by the 2013-2014 school year. These new salary schedules are based on effectiveness, demand of area of certification (which may include advanced degree levels), and experience, although the amount of pay raise was left entirely up to each district. The majority of current district salary schedules are based on level of education and experience alone, with experience outweighing degree attainment.

In March of 2013, a Baton Rouge judge ruled Act 1 unconstitutional because of its large size, citing that it contained too many items of legislation and violated the “single-object” rule of the state constitution. Following this decision, the Louisiana Supreme Court reviewed the ruling and asked the judge to reconsider in light of their ruling on Act 2, commonly known as the state’s voucher bill, that also was accused of violating the “single-object” rule, but was later ruled as constitutional. In August of 2013, a Monroe judge again ruled Act 1 unconstitutional, this time for violating a teacher’s right to due process.

The latest educational reform efforts in Louisiana include legislation that will eliminate teacher tenure and focus on teacher evaluation linked to pay for performance. These radical changes are occurring within a time of much more state legislation relating to both K-12 and higher education surfacing in a post-Katrina era when the state of Louisiana education is still facing recovery. These policies, however controversial, will leave lasting implications on the entire Louisiana education system. Each individual district within the state has reacted to the policy changes based on the unique cultures, histories, and people of each region

## **The School District**

The school district chosen for this study is located in southeast Louisiana, approximately fifty miles south of the city of New Orleans. The parish has a population of approximately 96,000 people, with slightly over fifteen percent of the population falling below the poverty line (United States Census Bureau). The district is known for ample wildlife for hunting and fishing as well as a port on the Gulf of Mexico that is responsible for furnishing the country with about 18% of its oil supply. Its rich cultural history as a land settled by the Cajuns has led to the development and retention of several small, tight-knit communities along the bayou that runs the length of the parish.

Students of the district can attend schools in the regular public school system, one of the three charter schools in the district (one of which is a virtual school), or one of the six Catholic schools (five pre-k-7<sup>th</sup> and one high school). In the public school system, the publicly elected school board establishes policies for the 15,000 students in 30 schools staffed with approximately 2,300 employees.

The 2012-2013 COMPASS report (LDOE, 2013) compiled data on each individual district's student, teacher, leader, and counselor scores. According to this report, 70.5% of students in this district were on grade level, compared to 68% of students across the state; 91% of teachers were rated Effective: Proficient or Highly Effective in their final COMPASS rating, compared to 89% of teachers across the state, although the Professional Practice rating (qualitative portion of COMPASS) showed 90% of teachers at Effective: Proficient or Highly Effective and the Value-added rating (quantitative portion of COMPASS) calculated only 60% of teachers to be Effective: Proficient or Highly Effective. This reveals a



disproportional result between the classroom observations conducted by principals and the value-added scores given to teachers based on student growth on standardized tests or on Student Learning Target (SLT) goal attainment for those Non-Tested Grades and Subjects (NTGS).

### **Review of Literature**

In order to fully comprehend the COMPASS evaluation system and its effects, it is important to gather a range of studies that have examined teacher evaluation. The definition of teacher evaluation has shifted considerably over the past several decades from a checklist of performance standards and responsibilities to standards-based principal observations and student achievement data measurements. The stated purpose for teacher evaluation has generally been to increase student learning by informing teaching practices. The COMPASS teacher evaluation system includes a combination of qualitative data from principal observations using a standards-based rubric and quantitative data from value-added scores or Student Learning Target achievement. Because the quantitative portion of COMPASS, according to law, can be used to challenge the qualitative scoring by the principal, it is vital to understand the research behind value-added measurements. The following review of literature will outline research on the history of teacher evaluation, the validity of standards-based rubrics for principal observations, the research on value-added models, and the principals' role within teacher evaluation.

### **History of Teacher Evaluation**

Teacher evaluation can be traced back to 1913 when Joseph Taylor first created rating scales to judge teachers' influence on students, teaching ability, enthusiasm, discipline, and energy (Callahan, 1962). The earliest studies of teacher evaluation reveal

problems with teacher evaluation systems that still appear in much more recent literature, such as whether or not teacher evaluation systems are serving their intended purpose of informing instructional practice. Early evaluation systems were quite different from the systems of today, focusing on principal observation as the main and sometimes only source of data, and not including any widespread best practice criteria as standards. In many cases, teacher evaluations included up to 75 percent of supervisory criteria (adherence to school policy, appearance, personal relationships, relationships with parents and the community, etc.) with little attention paid to instructional practices (McGreal, 1982).

After working with over 300 school districts on teacher evaluation, McGreal (1982) argued for complementary procedures within teacher evaluation systems that focused on instruction rather than administrative responsibilities and duties when the intended purpose of teacher evaluation is for increasing instructional practices. This article is one of the first to call for more sources of data to be included in teacher evaluations, including student achievement data. McGreal (1982) also identifies a major shortcoming of early teacher evaluation systems as a lack of training of both administrators and teachers on the system, leading to a “falling back” on old practices and attitudes (p. 305).

Lewis’s (1982) early study of educational personnel evaluation surveyed over 400 respondents for information about teacher evaluation. The data collection included checklists, outlines, and evaluation instruments used for teacher evaluation. Lewis (1982) found that the true challenge for school administrators was to successfully use personnel evaluation to improve teaching. This early study was the beginning of many that began to question the usefulness of teacher evaluation as a means to increase teacher effectiveness.

Darling-Hammond, Wise, and Pease (1983) reviewed the early literature on teacher evaluation and reported the four main purposes of teacher evaluation at the time were individual staff development, individual personnel decisions, school improvement, and school reputation. They discuss that individual and organizational needs were in competition through teacher evaluation and an ideal teacher evaluation system would find a way to satisfy both. They argue that research on teacher performance and teaching effectiveness does not lead to a stable list of measurable teaching behaviors effective in all teaching contexts and that research on individual and organizational behavior indicates the need for context-specific strategies for improving teaching rather than system-wide hierarchical efforts, which is what was eventually created through standards-based teacher evaluations.

From these early studies, teacher evaluation became increasingly thorough. *A Nation at Risk* (1983) originally called for district use of teacher evaluations for salary, promotion, tenure, and retention decisions. Wise et al. (1985) was one of the first empirical research studies to examine teacher evaluation in the context it is used today. Several questions are posed about the use of teacher evaluation that are still salient today. This case study examined four school districts that used teacher evaluation for personnel decisions and staff development. The findings of this study indicate that principals should be relieved of some managerial tasks in order to have adequate time for more powerful observations and conferences, master teachers should be used to strengthen the evaluation process and give support to teachers, and teacher evaluation should be a district commitment that formulates specificities of the system to meet individual goals without state-imposed highly prescriptive teacher evaluation systems (Wise et al., 1985).

In 1996 a seminal work that has had lasting impact was first published. Charlotte Danielson's *Enhancing Professional Practice: A Framework for Teaching* was a comprehensive model that sought to honor the complexity that is classroom teaching. She named 76 elements of teaching, broken down into four levels of performance. Hers was the first model to include the knowledge, skills, and dispositions of classroom teaching along with elements of supervision of classroom instruction. Soon, schools and school districts were using the framework as a guide to coach teachers (Marzano, Frontier, & Livingston, 2011, p. 54).

Originally, teacher evaluations were similar across the nation, a pass/fail checklist evaluation conducted once or twice a year by the school leader that the vast majority of teachers passed. Little attention was usually paid to results of these evaluations (Wise, et al., 1985). George W. Bush's education reform plan, *No Child Left Behind* (2001), brought to the forefront a focus on high-stakes standardized testing of students to rate student achievement and the "highly qualified" status of teachers which valued teacher knowledge of and experience in their content area(s) because it required a certain number of credit hours within subjects (Ahn and Vigdor, 2013). Although critics of high-stakes testing demanded that the nation's students were becoming overly tested through substantial time being spent on standardized testing, the emphasis on high-stakes testing continued in the Obama administration (Ahn and Vigdor, 2013; Kohn, 2000).

Despite *No Child Left Behind's* focus on teacher evaluation, classroom observations as part of evaluations were still criticized. A 2009 study examined twelve school district across four state and found that *The Widget Effect* (Weisberg, Sexton, Mulhern, & Keeling, 2009) "describes the tendency of school districts to assume classroom effectiveness is the

same from teacher to teacher” (p. 4). This study found alarming flaws in teacher evaluation systems: “73 percent of teachers surveyed said their most recent evaluation did not identify any development areas, and only 45 percent of teachers who did have development areas identified said they received useful support to improve” (p. 6). The Widget Effect blamed teacher evaluations that were “short and infrequent” and “conducted by untrained administrators” for a failure to reward effective teachers and remediate struggling and mediocre teachers. New federal policies would seek to mend what had become known to many as a meaningless and broken system.

*Race to the Top* (2009) gave grant funds to states that applied for innovative projects to improve education. A stipulation of this grant funding was for states to use student achievement data within its teacher evaluation systems, referred to as value-added measures (United States Department of Education, 2009). Many states, including Louisiana, passed state laws to include value-added measurements as a part of teacher evaluation systems and retained these policies even though they were not awarded the federal funding.

### **Value-Added Measurement: Quantitative Teacher Evaluation**

Value-added models measure the gain in a student’s scores during a certain period of time. Critics of subjective principal evaluation systems have turned to value-added models as the answer to determining teacher effectiveness through the use of student achievement as evidence. However, value-added models have their own critics who argue about the validity of the standardized tests themselves as well as the statistical measures used to calculate value-added measurements.

California was the first state to use a pre-cursor to what we know today as value-added models (VAM). In 1971 the California Education Code, known as the Stull Act, was signed into effect, forcing school districts to use student achievement as a portion of a teacher's evaluation. However, what student data used was and still is not prescribed by the Stull Act, so a variety of student achievement data is used (California Department of Education, 2012).

Tennessee was the first state to calculate value-added measurements as we know it today for its teachers using the Tennessee Value Added Assessment System (TVAAS) created by Dr. William Sanders (Tennessee Year One Report). Wright, Horn, and Sanders (1997) used a mixed-model analysis of variance using data from the 1994 and 1995 TCAP scores to examine whether or not teachers make a difference on student achievement. TCAP tests are given each spring to all students in Tennessee in grades two through eight. This study used combinations of TCAP scores and Tennessee Value-Added Assessment System (TVAAS) scores to examine the relative magnitude of teacher effects on student achievement while simultaneously considering the influences of intraclassroom heterogeneity, student achievement level, and class size on academic growth. The study found that teacher effects are dominant factors affecting student academic gain and that the classroom context variables of heterogeneity among students and class sizes have relatively little influence on student achievement. "Effective teachers appear to be effective with students of all achievement levels, regardless of the level of heterogeneity in their classrooms" (p. 63). A conclusion from this study argues for the use of student achievement data to evaluate teachers: "These results suggest that teacher evaluation processes should include, as a major component, a reliable and valid measure of a teacher's effect on student

academic growth over time. The use of student achievement data from an appropriately drawn standardized testing program administered longitudinally and appropriately analyzed can fulfill these requirements” (p. 66). This study claims that its findings prove that teachers do make a difference in student achievement.

In response to Dr. William Sander’s value-added model measurements, critics began to outline what they argued were more effective measurement tools. Alicia (2005) examined the TVAAS as created by Dr. William Sanders for Tennessee and asserts that the value-added model “appears flawed essentially because it assumes that the gain score of students is attributable only to the teacher(s)” (p. 1). Alicia (2005) argues for a value-added measurement that accounts for different student variables like socio-economic status and IQ. But those who are opposed to different calculations of value-added measurements also argue that the standardized tests themselves that are used to gather the data are not reliable indicators of student achievement because they are not valid assessments of student learning from year to year (Ahn and Vigdor, 2013). Fuller’s (2006) study of one thousand teachers and parents in Texas found that less than 10 percent of teachers and less than 30 percent of parents agreed that the Texas’s statewide high-stakes standardized test accurately assessed a student’s academic level. In many states, the selection of which student data to use has caused problems in the reformation of teacher evaluation (Ovando & Ramirez, 2007). The change to Common Core State Standards and assessments based on these standards has increased the challenge in using student test data to determine a teacher’s value-added score in Louisiana. Value-added measurements in many states, including Louisiana, control for external factors but still draw much

criticism from educators and testing experts (Amrein-Beardsely, 2008; Baker, et. al, 2010; Kersting, Chen, & Stigler, 2013).

Stumbo and McWalters (2011) contest that standardized exams usually do not contain enough data to accurately determine a teacher's impact on student achievement. There is also argument about use of value-added measurements for the grades and subjects that do not have standardized tests (Stumbo and McWalters, 2011). Louisiana has adopted the use of teacher-constructed Student Learning Targets to quell this issue. Student Learning Targets (SLTs) are goals set by teachers and approved by administrators at the beginning of the school year based on pre-test data on state-approved assessments. At the end of the year, teachers must have grown their students to their specified goal in their SLTs. Because of the shift to Common Core State Standards-aligned assessments, the state of Louisiana has decided to use SLT data in place of value-added data for at least the 2013-2014 school year as the quantitative portion of teacher evaluation.

The earliest value-added measurements were used for research purposes, and now because *Race to the Top* calls for student achievement data as a part of teacher evaluation system, more and more states are integrating value-added measurements into teacher evaluation programs. In order to fully understand the implications of COMPASS, it is necessary to comprehend value-added measurements because it not only is fifty percent of the teacher evaluation, but it also can be used to refute the qualitative half of the score, according to Act 1. Fuller and Hollingworth (2013) recently examined the use of VAM scores to assess principal effectiveness and concluded that "policy-makers (should) not use statistical estimates of principal effectiveness to judge, rate, or evaluate principals in any high-stakes manner" (p. 28). Instead, they argue for the use of VAM to help districts decide



where to focus attention rather than as a rating for principals or individual teachers. As evidenced above, policy-makers have questioned the validity of using standards-based teacher evaluation systems that could be subjective in nature, and in answer have called for studies like the following to examine correlations between principal observations using standards-based rubrics and value-added measurements.

### **Standards-Based Teacher Evaluation Systems: Qualitative Teacher Evaluation**

Although the earliest teacher evaluation systems did not utilize standards-based evaluations to assess teacher competency, Charlotte Danielson's original *Framework for Teaching* (1996) shifted many school districts' assessment criteria to include either portions of or the entirety of her identified elements within four domains of effective teaching characteristics. Danielson has revised the *Framework* several times, the latest of which is a 2013 edition. The *Framework's* intended use is "the foundation of a school or district's mentoring, coaching, professional development, and teacher evaluation processes, thus linking all those activities together and helping teachers become more thoughtful practitioners" (The Danielson Group, 2011, para. 3). Several studies have examined the relationship between student achievement data and use of the *Framework* in principal observations in an effort to prove some type of correlation between the two.

Validity studies of the *Framework for Teaching* have primarily focused on the correlation between teachers' evaluation ratings and teacher's effect on student achievement, also known as value-added measurements. Milanowski, Kimball, and White (2004) studied the relationship of evaluation ratings from *Framework for Teaching* -based systems and value-added measurements in three different school districts over a three-year span and found positive correlations. Milanowski, Kimball, and White (2004) argue

that the relationship between standards-based teacher evaluation scores and measures of student achievement needs to be demonstrated before using these scores in research on teacher effects or teacher quality. Milanowski, Kimball, and White (2004) continue research in these three sites in this study by extending the time period and sample size to find out whether the relationships they first calculated is consistent from year to year. Their analysis was based on the value-added paradigm, using two-level hierarchical linear models in which individual student achievement on a subject test is represented as a function of the prior year's test score in that subject, and a variety of student-level control variables intended to represent factors associated with test performance but that are not in teachers' control, such as ethnicity and English proficiency. The methods also included an analysis of correlations between teacher evaluation scores and estimates of average student achievement without controlling for student characteristics and found that there is little difference between the correlations with and without the controls. The results reported show that the scores produced by these standards-based teacher evaluation systems have a substantial positive relationship with the achievement of the evaluated teachers' students. The results are comparable with the earlier results of analysis of these sites, so Milanowski, Kimball, and White consider the results to represent a constructive replication of their earlier results.

Similarly, Kane, Taylor, Tyler, and Wooten (2010) also studied the validity of a Framework for Teaching-based evaluation system and found that teachers in the top value-added quartile consistently received higher ratings on all the standards in the prior year than those on the bottom, demonstrating that teachers who scored higher on their evaluations did indeed have more student learning growth.

Schacter and Thum (2004) and Daley and Kim (2010) explored the validity of another evaluation system that is based in part on the *Framework for Teaching*, the National Institute for Excellence in Teaching's (NIET) Teacher Advancement Program (TAP) model. The rubrics used in these evaluations are more specific about practices relating to student engagement, teaching to standards, higher order thinking skills, use of assessment, and differentiation of instruction, although they do roughly parallel the *Framework for Teaching* rubrics. The TAP evaluation system includes multiple observations by school administrators, master teachers, and mentor teachers to factor into a determination of teacher bonuses. Schacter and Thum (2004) used evaluations on 52 teachers by trained outside evaluators using the TAP rubrics on eight classroom observations of each teacher and found correlations of ratings with classroom value-added of .55 to .70. This high correlation proves that the teacher's value-added scores and the observations using the TAP rubric by outside evaluators were aligned. NIET's Daley and Kim (2010) used a larger sample of 1,780 teachers and found a correlation coefficient of .21, which causes the earlier study of 52 teachers to be less credible. Milanowski (2011) in a review of validity of *Framework for Teaching*-based teacher evaluation models research posits that perhaps the amount of specificity within the TAP rubric caused the high correlation values in these two studies.

The research on standards-based evaluation systems has mostly compared evaluation scores and value added scores, even though the validity of each has been questioned. Researchers have used value added scores to validate standards-based evaluation rubrics and in turn have used evaluation rubrics to assess the validity of value-

added scores. Another compounding issue of teacher evaluation is the principal's complex role as evaluator and supporter.

### **The Principal's Role in Teacher Evaluation: Problems and Issues**

The principal's complex role as teacher evaluator has placed increased pressures on administrators. Principals and teachers have identified several problematic areas within teacher evaluations as it pertains to principals as evaluators. It has been discovered that principals can be unwilling to score teachers too low, and that principals find it difficult to tier mid-level teachers who are not necessarily high or low performing. Teachers outline concerns about principals as evaluators as having a lack of subject area knowledge, failure to provide feedback and inconsistency among evaluators (Milanowski and Heneman, 2001). This concern with inter-rater reliability is present across several studies (Milanowski and Heneman, 2001; Donaldson and Papay, 2012; Canelake, 2012). The inability of evaluators to score low performing teachers and distinguish middle-tiered teachers can be tied to a lack of training as another issue that has been exposed through research (Donaldson, 2009; Canelake, 2012).

Tennessee has been a recognized leader in calculating value-added scores for teachers, with their first use of VAM dating back more than twenty years. In 2012, Tennessee implemented an evaluation system based on 50 percent quantitative scoring and 50 percent qualitative evaluation using principal observations with a rubric based on National Institute for Excellence in Teaching's evaluation model. The Year One Report released by the state education department revealed that evaluators, administrators, did an effective job identifying their higher performing teachers when comparing VAM scores to

observation scores; however, teachers with the lowest VAM scores were still rated as average by evaluators, “demonstrating an inability or unwillingness on the part of evaluators to identify the lowest performing teachers” (Tennessee Year One Report, 2012).

Principals play a critical role in the teacher evaluation process; however, Jacob and Lefgren (2008) found that “one should not rely on principals for fine grained performance determinations as might be required under certain merit pay policies” (p. 129) because principals could not be reliable evaluators of teacher performance. The study collected student and teacher data from a mid-sized school district in the western United States. The student data included demographic variables as well as standardized achievement scores, and the teacher data is linked to students and included a variety of teacher characteristics. The researchers also surveyed all elementary school principals and asked them to rate the teachers in their schools along a variety of performance dimensions. In general, the principals rated the teachers quite high across the board. The researchers used their own value-added measurement to measure student growth. The findings of this study suggest that principals can identify the best and worst teachers at their schools, but they have difficulty distinguishing the teachers in the middle.

Milanowski and Heneman’s (2001) study reports on teachers’ reactions to a pilot implementation of a new standards-based teacher evaluation system based on the Framework for Teaching in a medium-sized Midwestern school district. Most teachers interviewed and surveyed accepted the evaluation standards and the need for an evaluation system but perceived that the system added more to their workloads and did not provide enough feedback. Milanowski & Heneman (2001) identify three areas of

concern that teachers have with principals serving as evaluators: lack of subject matter knowledge, failure to provide helpful feedback, and inconsistency among evaluators.

The National Education Association (2010) reports that the majority of principals have not been trained in evaluation methods to the degree at which they can provide feedback to teachers. Teachers will only use feedback from administrators when they believe their evaluator is skilled and competent. Donaldson (2009) examined current teacher evaluation programs in an effort to discover why teacher evaluation has little effect on instruction, learning, and achievement. She discovered a lack of training of evaluators and argued that administrators must commit to training because “without high-quality professional development, evaluators will not evaluate accurately and the evaluation will likely have little impact on teaching or learning” (p. 11).

Donaldson and Donaldson (2012) outlined steps for strengthening teacher evaluation. They reported that a lack of trust and a lack of pedagogical knowledge inhibited the evaluation process from being a learning experience for teachers. Teachers need to trust their evaluators in order to “respond to them with efforts to build on their strengths and address their weaknesses, they must trust the observer and have access to subsequent learning opportunities” (Donaldson & Donaldson, 2012, p. 80).

Inter-rater reliability also comes into question as school leaders may all view the same evidence to rate teachers differently. Danielson (2011) concurs that “even after training, most observers require multiple opportunities to practice...to calibrate their judgments with others” (p. 38). Donaldson and Papay’s (2012) case study charted one district’s use of a teacher evaluation system from conception to implementation. The system was developed through a collaborative effort through the district’s 2009 collective

bargaining agreement with its teachers' union. It bases scores for teachers on three components: student growth on performance goals set by teachers and administrators, standards-based observations, and professional conduct. The results of this case study found that most of the participants viewed the district's teacher evaluation program positively. Many noted that the evaluation system still needed work. "Many participants said that there needed to be more standardization of the process and the ratings among administrators" (p. 41). This study calls for a calibration in observation and student-learning objective goal setting.

Teachers must believe their school leaders have sufficient training, a vast knowledge of both content and pedagogy, and have worked on inter-rater reliability with other evaluators in the school building and district in order to trust their evaluator when given feedback and opportunities for growth. Without these pieces, the evaluation process becomes nothing more than the assignment of a score.

### **The Principal's Role in Teacher Evaluation: Principals' Perceptions**

Research has also been conducted on principals' and teachers' feelings and perceptions of teacher evaluation. This is a critical part of understanding the implications of implementing teacher evaluation on the school level. Kersten and Israel (2005) found in their survey of 63 administrators that school leaders believe that the current teacher evaluation systems are much too time intensive and preclude many other opportunities for school building leaders to work with faculty to improve classroom instruction.

Because school leaders' time is so valuable, principals also report that they became frustrated when trying to learn a new rubric. Canelake's (2012) mixed methods dissertation explored the perceptions of nine school administrators through surveys and

interviews as they learned and implemented a teacher evaluation rubric based on the *Framework for Teaching*. The principals reported that as they used the rubric, they became increasingly frustrated with the instrument because of the difficulty to align it with certain instructional practices. They had difficulty ensuring inter-rater reliability across school sites and shared their struggles with using the teacher evaluation to actually improve professional practice.

Another issue discovered is a lack of alignment between the purpose of teacher evaluation and the actual implementation. Halverson, Kelley, and Kimball's (2004) study attempts to fill a void in the literature for research on how school leaders use teacher evaluation to shape teaching practices in schools. They examined through a case study a large school district that adopted the Framework for Teaching to evaluate teachers. Interviews of district leaders, principals, and teachers found that even though the stated purpose of the teacher evaluations by the school district was to impact student learning, principals and teachers did not see a direct correlation from teacher evaluation to student achievement. If the purpose of teacher evaluation is to increase teacher effectiveness and thereby impact student achievement, school leaders in their role as teacher evaluator should be able to recognize the connection from teacher evaluation to student achievement.

### **The Principal's Role in Teacher Evaluation: Instructional Leaders**

Limited research has been conducted that has explored principals' use of instructional leadership behaviors and actions in their role as evaluators. Ovando and Ramirez's (2007) multi-case study examines principals' instructional leadership actions within a comprehensive teacher evaluation system. The sample size was six school



administrators within the same school district. Three of the participants were principals and three assistant principals, representing elementary, middle, and high school. The data sources included interviews, observations, and journaling. Three instructional leadership actions were identified through inductive analysis as emergent themes:

1. setting clear expectations
2. monitoring instruction through walk-through observations
3. providing professional development opportunities according to teachers' needs

The findings of this study suggest that principals used the evaluation system to enhance instruction and improve student achievement, which is a stated purpose of the teacher evaluation system. These findings, however, cannot necessarily be generalized to a larger population because of the small sample size and because the schools that participated in the study had been identified as “successful” schools.

Sartain, Stoelinga, and Brown (2011) examined schools at various performance levels. They performed a two-year study of Chicago's Excellence in Teaching pilot in the eight elementary schools chosen to participate in the pilot, an evaluation program designed to give teachers evidence-based feedback on their strengths and weaknesses as classroom instructors using Charlotte Danielson's *Framework for Teaching*. This broad study examined the validity of students' growth data in relation to teacher observations and investigated the reliability of the observations by comparing multiple observations on the same teachers by different administrators. This mixed-methods study also gathered data from principal and teacher conferences and interviews with principals about their perceptions of the new evaluation system.

Sartain, Stoelinga, and Brown found that classroom observation ratings were valid measures of teaching practice when they were both compared and that the classroom observation ratings were reliable measures of teaching practice because evaluators watching the same lesson consistently gave the same ratings. Principals and teachers said that conferences were more reflective and objective than in the past and were focused on instructional practice and improvement; however, teachers complained that principals lack the instructional coaching skills to have deep and meaningful conversations about improving instruction.

Ovando and Ramirez's (2007) study found that principals were able to use teacher evaluation to support teachers and enhance professional growth; however, Sartain, Stoelinga, and Brown (2011) found that even though classroom observations were reliable measures of teaching practice, teachers did not feel that principals were able to provide feedback and coaching through the evaluation process that impacted their teaching.

### **Theoretical Framework**

Schools are open systems that are impacted by policy changes. Contingency theory's contingencies of leader behavior, leader-member relations, task structure, and position power, can be used to analyze principals' navigation of their role as teacher evaluator. Instructional leadership theory, a contingency model of leadership, can be used along with contingency theory to evaluate the behaviors of principals as they evaluate teachers.

### **An Open System**

In an era of school reform, public school education is facing change in nearly every facet: accountability measures and grading systems for schools, school choice, high-stakes evaluations for principals and teachers, student test data used for teacher pay-for-

performance and tenure, and increased rigor through a new standards system. Policy changes in education are external factors that create huge departures from past practices for educators at every level. Because of this, it is important to consider schools as open systems that are affected by environmental issues, one of those being policy changes.

The idea of social systems as open systems stems from the work of Max Weber (1947), although it was not until Talcott Parsons (1960), that the environment was recognized as a critical contributor to and a dependent of the organization. Contingency theory, an open systems theory, specifically examines how to lead an organization in a system that is affected by the environment and organizational variables (Marion, 2002). Schools are open systems because they are directly impacted by their environment, and they, in turn, affect their environments; therefore, school leaders can turn to contingency theory as a theory by which to draw practical implications.

### **Contingency Theory**

Fred Fiedler (1973), a pioneer in the work of contingency theory, developed three contingencies of appropriate leader behavior including *leader-member relations*, *task structure*, and *position power*. Because this study's focus is on the relationship between principals and teachers and principals' navigation of these relationships, these three contingencies can be examined to determine the degree to which these contingencies impact these relationships.

*Leader-member relations* refers to the relationships between the leader and group members. A leader who is accepted by the group is considered to be in a more favorable situation than one who is not. In schools, *leader-member relations* implies that principals and teachers may have pre-existing interpersonal relationships as well as professional

relationships. Especially in small-town or tight-knit community settings, as is the case in this study, these relationships could affect or be affected by high-stakes teacher evaluations. Principals who are in favorable leader-member relations situations may be considered more trustworthy and may have group members, teachers, who view their evaluations more positively or trust the results more because of this.

The *task structure* of an organization depend of several factors: if a decision can be demonstrated as correct, if the task is understood by all group members, if there is more than one way to accomplish the task, and if there is more than one possible solution to the task. In order for the task structure to be in a favorable situation, the group leader needs to have more knowledge about the task than the group members and the task must be structured. In the case of teacher evaluation, the greater the competency exhibited by principals as evaluators, the more teachers will trust their results and use results to improve instruction. Also, teachers need to understand the process (task) of their evaluations in order for this situation to be favorable.

*Position power* refers to the amount of power that the leader wields in relation to the group members over the specified task. With teacher evaluations, principals do have a great amount of power because the evaluations they conduct are used for personnel and merit-based pay decisions. In *position power*, the greater the power of the leader, the more favorable the situation.

Hersey and Blanchard's (1993) contingency theory model of leadership, a "situational model," named two main leader behaviors that appear as a result of Fiedler's three contingencies. *Task behavior* is to what extent the leader directs the individual on how to perform the task. It will be important to identify which behaviors of the principals

in this study fall into the category of task behaviors and how teachers react to this behavior. This is crucial because an influence on principal/teacher relationships will undoubtedly occur if the teacher feels the principal is providing the necessary information and direction in order for them to be successful on a COMPASS evaluation. The opposite of this could also possibly be true in that teachers could question the expertise of principals to evaluate their classroom practices. *Relationship behavior* is the extent to which the leader communicates, usually in a way that is nurturing, supportive, and encouraging. These behaviors will also need to be identified along with teachers' reactions about the support or lack of support they feel. Because COMPASS evaluations take considerably more time than previous evaluation systems, principals may be able to support teachers less than before, or teachers could possibly feel a more specific rubric is giving them increased administrative support.

### **Instructional Leadership**

Instructional leadership theory is one of several contingency models of leadership; it emphasizes the importance of leaders' heavy involvement in teaching and learning (Hoy & Miskel, 2008). This model for leadership surfaced in the 1980s in research on effective elementary schools in poor urban communities (Edmonds, 1979; Liethwood & Montgomery, 1982), and quickly gained support as policymakers urged principals to adopt this model as a means to increase school effectiveness (Barth, 1986). They were defined as hands-on principals who were entrenched in curriculum development and experts of instruction who worked directly with teachers to improve student achievement (Cuban, 1984; Hallinger, Bickman, & Davis, 1996; Hallinger & Murphy, 1986).

The instructional leadership model that has been used most often for empirical research (>125 studies) is the Hallinger and Murphy (1985) model. The model was developed by Hallinger and Murphy as they faced the task of evaluating the extent to which administrators in their district exhibited the behaviors of instructional leaders. Up to this point, no clear and explicit model had been developed, so they created this model and a rating scale. The *Principal Instructional Management Rating Scale* has been used for over two decades to evaluate principal behavior. The model from which the rating scale derives divides the roles and responsibilities of an instructional leader into three dimensions with ten descriptors. *Defining the School Mission, Managing the Instructional Program, Developing the School Learning Climate Program*. Although this model includes descriptors that are unarguably critical attributes of an instructional leader, it does not contain any mention of behaviors relating to the relationships of the instructional leader or any support systems for the teachers. At a time when educational policy changes are placing increased pressures on teachers, the relationships between principals and teachers have become vitally important as new forms of schooling have opened alternate career pathways for teachers. Teachers may now search for “greener pastures” when faced with more pressure and less support from their administrators.

Because the COMPASS teacher evaluation system is so different from systems of the past, Hallinger and Murphy (1985) did not realize that not including a leadership function of supporting teachers through their evaluations would be such a critically missing piece of their leadership model. It is now crucial for principals to support teachers before and after their evaluations. This gap can be filled with Hersey and Blanchard’s (1993) contingency theory model of *task behavior* and *relationship behavior*. If school leaders engage in task-

oriented behavior that enables a task that needs to be done in a way that supports and encourages teachers (*relationship behavior*), those who are uncomfortable with change (the COMPASS evaluation system) will gain confidence and become motivated. In a school setting, this would look like principals conducting walkthrough visits prior to a formal evaluation to provide teachers with feedback. Feedback would be given during face-to-face conferences. Teachers would be given appropriate professional development on the rating system being used for their evaluation and would meet with the principal prior to their formal evaluation to ensure transfer of understanding has taken place. Following the formal evaluation, principals conduct a post-conference in which teachers are provided with areas of strength and weakness and specific instruction on growth in the area of weakness is provided to the teacher.

### **Why These Theories**

It is more critical than ever that school leaders possess the qualities of instructional leaders now that they are being asked to conduct high-stakes evaluations of teachers. In order for these evaluations to be fair and accurate and to help produce the desired effects (teacher growth in effectiveness), school leaders must hold the knowledge of curriculum and instruction to be effective evaluators themselves. For this reason, examining behaviors of principals during the teacher evaluation process through the lens of instructional leadership is a relevant fit. It is proposed, however, that the most popular model of instructional leadership is somewhat antiquated and needs an additional dimension to be an appropriate measure of instructional leadership behaviors. This gap can be filled with Hersey and Blanchard's (1993) contingency theory model of *task behavior* and *relationship behavior*. This study will examine the stories of school leaders as they make meaning of

teacher evaluation policy and how they perceive their roles as they undergo the process of evaluating teachers in their school building.



## **Chapter 3: Methods**

### **Personal Reflection: My I as an Evaluator**

At the end of my fourth year as a high school English teacher, I received a phone call from a friend asking me to apply as a TAP Mentor Teacher at a local middle school. The school was closer to my home, and with a newborn at home and the promise of a partial teaching load, I applied. I hesitated in accepting the position when it was offered. A large part of me did not feel ready to leave the classroom, my beloved Advanced Placement English students, and a school faculty I adored. Because I felt it was best for my family, I joined the middle school world and became essentially a part time teaching coach and part time interventionist for at-risk students. My I as an evaluator emerged here as I trained to be a TAP evaluator. The training was extensive, and an assessment followed to test my ability to score and coach teachers using the extensive TAP rubric. I loved everything about the TAP rubric that I did not like about the COMPASS rubric used to evaluate me as a teacher the year before. I still feel as if the TAP rubric is more comprehensive and gives a more complete picture of a teacher's effectiveness. I also experienced being on the other side of teacher evaluation. As an evaluator, I scored and coached teachers who were double my age, out of my content area, first-year uncertified teachers, and most difficult of all, some I considered my friends as well as colleagues. Straddling the line between evaluator and teacher was a difficult role, but it was one that offered infinite learning opportunities. My I as an evaluator is two-fold: I used a different, and in my opinion, preferable model to evaluate teachers, and I also experienced the side of evaluating teachers. Throughout this study, I have had to detach myself, especially when participants chose to discuss previous TAP experience. I was careful to allow their opinions of COMPASS, and not necessarily my

own, be voiced. I could also identify with much of their stories as they discussed navigating coaching and evaluating. When you work closely with improving a teacher's practice then evaluate them, you root for them to succeed, and it can become difficult to not allow this to influence your interpretation of their lesson for an evaluation.

### **Introduction**

This chapter revisits the purpose of this study and the research question explored. It also discusses in-depth the research design, data collection, data analysis, and my own role as the researcher. It includes information on the ethical considerations and how the methods of this study were verified.

### **Purpose**

The purpose of this study is to explore school leaders' stories about their roles as high stakes evaluators of teachers. This narrative study uses stories to garner school leaders' integration of personal abilities, role conception, and understanding of the policy. The use of principals' stories to explore this information will increase understanding of the roles principals must now play as the instructional leader of schools through the use of teacher evaluation. This strives to impact the practice of school leaders as well as give further implications for the policy development of teacher evaluation.

### **Research Question**

The goal of my study is to examine the attributes, perceptions, and actions of school leaders as they approach the process of high stakes evaluations. The following main research question is explored by this study: *How do principals describe their role as high-stakes evaluator through the implementation of COMPASS?* Through a narrative methodology, principals told their own stories through interviews.

## Research Design

Qualitative research begins with a theoretical framework by which researchers view a problem; “qualitative researchers use an emerging qualitative approach to inquiry, the collection of data in a natural setting sensitive to the people and places under study, and data analysis that is both inductive and deductive and establishes patterns or themes” (Creswell, 2013, p. 44). The study “includes the voices of participants, the reflexivity of the researcher, a complex description and interpretation of the problem, and its contribution to the literature or call to change” (Creswell, 2013, p. 44). This study’s theoretical framework uses an existing model of instructional leadership and Hersey and Blanchard’s (1993) contingency theory model of *task behavior* and *relationship behavior* (Hoy & Miskel, 2008). This study examines the stories of how school leaders perceive their roles as they undergo the process of evaluating teachers in their school building.

This study’s goal is to examine the stories of school leaders as they use high stakes evaluations of teachers. The findings from this study have implications for practitioners as they engage in similar situations and for policy-makers as they create new policies for teacher evaluation. Because qualitative research “includes the voices of participants,” it is the most appropriate research design for this particular study. Qualitative research also allows for “reflexivity of the researcher,” and because the researcher has been a teacher and an evaluator in the evaluation process, reflexivity was a vital component of the research process.

A narrative approach to inquiry “begins with the experiences as expressed in lived and told stories of individuals” (Creswell, 2013, p. 70). It is a qualitative design that is “understood as a spoken or written text giving an account of an event/action or series of

events/actions, chronologically connected” (Czarniawska, 2004, p. 17). Different forms of data, including interviews, observations, conversations, and documents are arranged in some type of chronology and can be analyzed through several different approaches (Creswell, 2013). According to Elliot (2005), narrative studies are chronological, meaningful, and social because they are arranged in sequential order for a set purpose and audience. A study with a narrative design collects the stories of individuals that are “co-constructed between the researcher and the participant” (p. 71); therefore, Creswell (2013) calls for a “strong collaborative feature of narrative research” (p. 71). Because the researcher is a former public school teacher still working in the public school system, the collaborative feature is appropriate for this study. Narrative stories “may shed light on the identities of individuals and how they see themselves” (Creswell, 2013, p. 71). Because of this, a narrative approach to discovering changes in school leaders’ behaviors through the evaluation process is exceptionally fitting. Through the discovery of these changes, information about how principals can appropriately and effectively navigate this role was gleaned as well as information for policy-makers about school-level implications of teacher evaluation policy.

### **Participants**

Seven participants were selected for this study. Four of the participants responded to an initial email requesting participation from secondary principals within the selected local school district, which will be referred to as Local School District (LSD). In an effort to recruit one more participant to reach the goal of six participants in the study, another email was sent to secondary assistant principals. Two assistant principals immediately responded as well as another principal, so the participant total reached seven. All of the

participants were white, between the ages of 31 and 60, had at least ten years experience in education, and had conducted at least 16 COMPASS teacher evaluations. All of the participants had prior experience with at least one type of teacher evaluation system other than COMPASS. All of the interviews took place in each participant's office. The interviews transpired during the months of April, May, and June 2014.

## **Data Collection**

### **Gaining Access**

I gained site access through approaching the gatekeeper, the district superintendent. Next, I contacted the school principals within the Local School District (LSD). One middle school principal was not contacted in the LSD because this school is a TAP (Teacher Advancement Program) school, so this school utilizes the TAP evaluation system instead of COMPASS. In addition to this, it was the school I worked at during the time of data collection, so it was not used for this study. My previous school leaders were not considered for the sampling because this leads to "biased data selection and decreased objectivity" (Glesne, 2011, p. 147). For this reason, I did not contact the school principal at the site I previously taught at, a high school in the district. The principals in the study are not principals that I have had anything more than brief contact with. I have professional but not personal relationships with the school leaders who participated in the study.

### **Sampling**

Following IRB and gatekeeper approval, I used purposeful sampling to contact the remaining secondary school administrators through email and asked for participation in my study. I received confirmation from seven potential participants initially. Two potential participants resigned from the study, and one reconsidered and did participate in the study

after I had located six participants, bringing the total number of participants to seven. This is one additional participant than the number of six that was originally intended. I then emailed each participant individually and explained the purpose for the study and clearly negotiated the responsibilities and expectations (Glesne, 2011).

### **Data Collection**

Each participant completed a brief demographic questionnaire requesting background information about their prior professional experiences as teachers and principals (Appendix B). The questions were planned prior to the first interview based on Chase's (2005) discussion of how to compose interview questions for narrative inquiry. The questions were broad, open-ended questions used to provoke story-telling and invite the participant's story (Chase, 2005). The recorded interview was transcribed within 48 hours of the interview. An additional interview followed with several additional questions about the participant's experiences with COMPASS.

### **Data Analysis**

#### **Narrative Analysis**

I used narrative analysis to analyze the interview transcripts. In narrative analysis, "the context in which the narrator tells the story influences what is told and how it is told" (Glesne, 2011, p. 186). I used the process of "restorying" to tie together the two interviews to form one coherent story of each participant. "Restorying" is the gathering of stories to analyze the key elements in a chronological order of events (Creswell, 2013). I first used chronological story mapping from the two interviews to piece together each participant's story. I sifted through the interviews and questionnaire and wrote a longer, chronological narrative of each participant. Each narrative began with the participant's first encounter

with COMPASS and ended with their most recent experience with it at the time of the interview. Some participants had provided a look ahead at what they were planning for the future of their schools with COMPASS. This gave me a complete picture of their experience with COMPASS, but it did not necessarily provide the most compelling story. In following narrative methodology, before I attempted a thematic analysis across participant's stories, I looked within each story for the voices of the participants to write a compelling story that strove to capture the essence of each participant (Chase, 2005).

From this initial chronological draft, I constructed a table that addressed how each participant answered my research question and the hallmark of each participant's story (Appendix D). This guided me as I composed story maps for each participant and worked to revise the narratives. I used story mapping to create plot diagrams of each participant's story, allowing each story to include an exposition, rising action, climax, falling action, and conclusion (Appendix E). Deciding which parts of stories to use was "an interpretive decision shaped...by theoretical interests" (Riessman, 2000). The exposition included some type of beginning with COMPASS even if it was not the participant's very first encounter, and the climax was selected from the chronological stories as the most interesting and unique experience each participant had with COMPASS. The actions leading to this climax, including some sort of conflict, whether internal or external, was selected for the story map. The actions following the climax and conclusion were also included in the story map. This whittled down the chronological stories into much more concise and interesting stories that strive to capture the voices of the participants as they view their roles in COMPASS. With my knowledge of the theoretical lens for this study, I also selected the plot based on the portions of the stories I knew would explore my research question. I used the

plot diagrams to restructure the initial chronological stories into a more concise and directed narrative following one specific plot line, which in each case is the one that I deemed to be the most compelling part or parts of that participant's chronological story. According to Riessman (2000), narrative analysis relies on detailed transcripts of interview excerpts. For this reason, in some instances the participants' own voices are used to tell a portion of their story. In other cases, their words are interwoven within my own interpretations of their stories. After completing the story maps, I rewrote each narrative following the plot diagrams.

### **Thematic Analysis**

Through thematic analysis, themes and patterns were located and analyzed. Comparisons were made between the different data collected, and unifying aspects of the data were located to help "reveal underlying complexities" (Glesne, 2011, p. 188). I coded the transcripts using coding schemes to help sort data and begin the analysis process. This began with rudimentary themes from the theoretical framework of this study and evolved into more complex coding as categories divided (Glesne, 2011). I looked for themes to emerge that clarified or explicated the roles of principals, looking especially at those that could be classified as task behaviors and relationship behaviors from Hersey and Blanchard's (1993) contingency theory model of leadership and placed throughout Hallinger and Murphy's (1985) instructional leadership theory model.

### **Verification**

#### **Member Checking**

I used member checking to verify that I have transcribed, interpreted, and analyzed the data I collected accurately. After transcribing the interviews and analyzing the data, I



sent the participants their own information that was used in the study (Creswell, 2013). By using this method, the researcher and the researched can grow in their interpretations of their experiences and stories (Glesne, 2011).

### **Peer Debriefing**

Throughout the process of data collection and analysis, a fellow doctoral student reviewed the data I collected and analyzed along with my own reflective journaling. My peer provided me with feedback about my methods employed in order to ensure that I articulated my participants' stories correctly and avoided subjectivity within my analysis. This kept the researcher honest about methods and the researcher's biases clarified through this interaction (Creswell, 2013).

### **Ethical Considerations**

The Institutional Review Board (IRB) approved this study before the collection of any data to ensure that the participants were respected as human subjects (Glesne, 2011). My participants were asked to sign letters of consent that clearly outlined the manner in which data was collected and used prior to submitting any information for data collection (Glesne, 2011). Pseudonyms as well as changes in gender were used in place of participant names in order to ensure a measure of confidentiality (Glesne, 2011). I describe only basic demographic information on the participant and the school site in order to help keep their anonymity and only the entire school district will be described in the context of the study in order to provide my audience with a background for the study. My participants were all informed of the purpose of my research in detail.

## **The Researcher's Role**

### **Field Relations**

Field relations begin with establishing rapport with the participants and leading that rapport into a relationship that includes trust. Participants were more willing to respond openly and honestly once this type of relationship has been constructed (Glesne, 2011). This was accomplished initially through email and then through our face-to-face meetings as I entered their school sites for my observations. I have had some prior contact with my participants through district meetings or even personal acquaintances that are not close in nature, but out of the seven possible participants for my study, I have not worked for any nor do I have any close personal or professional relationships.

### **Reflexivity**

Reflexivity "is an awareness of the self in the situation of action and of the role of the self in constructing that situation" (Bloor & Wood, 2006, p. 145). Researchers must control for subjectivity by tracing their own subjective selves through an awareness of our own identities (Glesne, 2011). One must be attuned to their own perspectives, opinions, and emotions throughout the data collection and analysis processes. This is especially important in a narrative study in which the research at times feels like they are 'running' alongside their participants as they experience their stories as they are told, retold, and in essence, relived (Etherington, 2004). I used the process of drafting personal reflections about my "subjective I's" throughout the data collection and analysis processes as a means to acknowledge "personal and theoretical attachments" and to allow for positionality of "personal factors that are either impossible or difficult to change" (Glesne, 2011, p. 157). As a former teacher who has undergone the same evaluation process that is being studied, it

would be impossible to not possess emotions and opinions about the topic, but through reflection, I acknowledged the subjectivity that occurred in the study.

### **Reflective Writing**

Subjectivity was an unavoidable component in the collection, analysis, and interpretation of data. As a former teacher who was evaluated by high stakes evaluations, I utilized an “interactive voice” to understand myself and the narrators (Chase, 2005, p. 666). I revealed my own “subjective I’s” involving teacher evaluation through my personal reflections throughout this study and thus became aware of how my own perceptions could be shaping my data analysis. Throughout the data collection and analysis processes I was aware of my own background and experiences and their ability to influence my interpretations of the data. As much as possible, I tried to reveal the voices of my participants in my analysis and interpret the data with as much objectivity as possible.

My purpose behind including reflective writing within my study is two-fold: I am interested in revealing myself to my audience because I feel that my own experiences as an educator can help to provide further insight into the context of the study, and I know that I am bringing with me my own subjective “I’s” that could lead my interpretation and I want to reveal my personal biases on this topic (Peshkin, 1988). My goal is to fully reveal the lens through which I am collecting and analyzing the data.

### **My Story**

I am a conventionally trained secondary English teacher who taught in a public high school for four years before transferring to a middle school as a TAP mentor teacher and reading interventionist. As a secondary English teacher, I helped to begin my school’s Advanced Placement program and served as the Advanced Placement Coordinator for a

year and a half. During my last year at the high school (the first year of COMPASS implementation), I underwent one announced COMPASS observation before leaving for the remainder of the year on maternity leave. During the semester before I left, I shadowed three assistant principals and accompanied them on five COMPASS evaluations of other teachers as part of an internship course to complete my leadership certification. When I was hired as a mentor teacher/interventionist at a TAP middle school, I attended an eight-day training on the TAP evaluation rubric and took a certification test to become a certified TAP evaluator. I evaluated teachers using the TAP evaluation rubric. As a mentor teacher, I was assigned eight teachers of various content areas to support in areas of refinement. I conducted walkthrough observations of all teachers and I used the TAP evaluation rubric for formal observations, pre-conferences, and post-conferences. I have never attended any training for the COMPASS evaluation system.

I hold a Master's of Education as a reading specialist, and I am certified to be an administrator in the state of Louisiana, although I have never been a school administrator. This year I accepted a position with a neighboring district as part of the curriculum department at a central office. I oversee the Kindergarten-12<sup>th</sup> grade English Language Arts curriculum. This includes composing district benchmarks, analyzing district data, leading choices about curricular materials, providing professional development to teachers and master teachers, and conducting walkthrough observations using some components observed on the COMPASS rubric. I do not directly lead any instruction specifically based on the COMPASS rubric, although I do frequently informally coach teachers about what evaluators look for in instruction during COMPASS evaluations.

## Chapter 4: Findings

The purpose of this study was to explore the stories of principals as they navigate their roles as high stakes evaluators of teachers. Principals' experiences with the COMPASS teacher evaluation system were explored over two interviews with each participant in an effort to collect data that would answer the question: *How do principals describe their roles as high-stakes evaluator through the implementation of COMPASS?* Principals told stories about COMPASS observations, their experiences in being introduced to COMPASS, and their interactions with teachers about COMPASS. Two main roles were identified as answers to the question through the lens of the Theoretical Framework that guides this study: *Instructional Coach* and *High Stakes Evaluator*. The participants generally seemed more comfortable in their role of instructional coach and less confident in the shift in their role from evaluator to high-stakes evaluator. They expressed frustration with what was revealed to be a low amount of *Position Power* as high-stakes evaluators (Fiedler, 1973). Even though they are principals and evaluators, the teacher evaluation policies and the COMPASS rubric itself can at times make principals feel like they actually have little control over the evaluation process, as evidenced in their narratives. The frequency, content, and structure of observations is now established by the state and participants perceived this as a loss of authority, even though their role as teacher evaluators ostensibly prioritized by state policy.

Chapter 4 begins with basic information about the participants and their schools. It continues with a narrative of each of the seven participants titled words that best describe their personally negotiated role as a COMPASS evaluator. Following this is a thematic analysis that divides the recurring themes from the data into two main roles the principals

exhibited during their work with COMPASS: *Instructional Coach* and *High Stakes Evaluator*.

The words of the participants were used in both the narratives and thematic analysis and give the participants a voice in the research. The participants' responses answer the research question and the framework by which the data is analyzed. Table 1 below provides a visual depiction of participants and the following information, which was gathered from the questionnaire the participants answered during Interview 1.

**Table 1: Profile of Participants**

	Age	Gender	Years in education	Years in admin role	Years in current position /Type of Role <sup>1</sup>	Total number of COMPASS evaluations conducted	Type of School
Participant A: Mr. Boudreaux	31-40	Male	14	7	2/P	34	Middle
Participant B: Mr. Breaux	41-50	Male	18	8	2/P	75	Middle
Participant C: Ms. Robichaux	41-50	Female	20	9	2/P	24	High
Participant D: Mr. Richard	51-60	Male	31	11	5/P	50	Middle
Participant E: Ms. Simoneaux	31-40	Female	10	2	1/AP	16	Middle
Participant F: Mr. Thibodeaux	31-40	Male	13	4	2/P	35	High
Participant G: Ms. Fuselier	31-40	Female	16	2	2/AP	60	High

---

<sup>1</sup> P= Principal, AP= Assistant Principal

## **Narrative Analysis: Stories of Principals as Evaluators**

All of the participants are natives of Southeastern Louisiana. Each of the participants in this study was given a pseudonym to represent them. The pseudonyms were chosen from Cajun French and Creole names that are popular in Southeastern Louisiana to honor the unique culture that is present in the LSD. Each participant's story is represented below from the information gathered from two interviews. Although the information was not told in the order it is presented, *restorying* was used to string the events into a logical, chronological order in an attempt to answer the study's research question through descriptions of the participant's experiences with COMPASS evaluation system (Creswell, 2013).

### **Mr. Boudreaux's Story: Trying to obtain an accurate picture of teaching effectiveness**

Mr. Boudreaux is the principal of a small middle school with approximately 300 students and a fifty percent free and reduced lunch rate. It is considered a community school, with the students of just a few small communities attending. The school was given a C rating by the Louisiana Department of Education for the 2013 school year. Mr. Boudreaux is one of two evaluators at his school, so he splits the COMPASS teacher evaluations with his assistant principal. He is in his thirties and has been the principal of the middle school for two years. Before this, he was an assistant principal in a neighboring district where he participated in the pilot program for COMPASS, which gives him a unique perspective of COMPASS.

Mr. Boudreaux was introduced to COMPASS during the pilot program three years ago, and the first part of learning the evaluation system was to read Charlotte Danielson's *Framework for Teaching*. Next came "about three days' worth of training on exactly how to

score given teachers on given attributes based on what they presented to us both in portfolios and interviews and classroom observations.” The pilot for COMPASS included a portfolio of each teacher’s observations and artifacts, and the observation rubric included the entire Charlotte Danielson *Framework for Teaching*. Mr. Boudreaux described it as “ a very...very...deliberate, complex process that took 45 minutes of pre-observation just to go through all the artifacts, then an observation that was the entire rubric, then a post observation which took at least another 45 minutes, and then throughout the year it was just assessing where they were on given attributes that were assigned as areas of needed growth.” After this came walkthrough observations and a final assessment. “It was a very, very in-depth process that was ultimately deemed by the state department too cumbersome and time-consuming for administrators to actually do.” After experiencing the process during the pilot, Mr. Boudreaux found his first observation using the five parts of what became the COMPASS rubric surprisingly easy.

Mr. Boudreaux applied for and received the job of principal at his current middle school in LSD. The pilot program was shortened into the current version of COMPASS: two formal observations, one announced and one unannounced, no portfolio, and shorter versions of the conferences. Mr. Boudreaux’s first observation with the five-section COMPASS rubric was “surprisingly easy” and it “went well” even though it was unannounced and he “could tell she was a little uncomfortable.” The difficulty came in when trying to score the observation on the five sections of the new rubric. Mr. Boudreaux looked for the “attributes of COMPASS, which, uh, are limited, and the rubric doesn’t align very well, either.” Within each category of COMPASS is a descriptor of teacher and student behavior. In some cases, the descriptors do not continue in different levels, but rather are



left off the rubric. For example, “The teacher makes effective use of wait time” is found under the Effective: Proficient range but is not mentioned under another achievement level. If a teacher is not using wait time, it can leave an evaluator wondering how to score for that area. Or does an evaluator not worry about wait time unless it is used effectively? Because of this, Mr. Boudreaux feels that “it’s not a rubric. What you have, in each of the components, are disjointed and not related to one another’s attributes...it was never designed to be a rubric of evaluation, so anybody that has read the book can clearly see that.” Trying to adapt to using the five part COMPASS rubric following his previous experience has been “a frustrating process to go through to some extent.” Despite his dissatisfaction with the rubric itself, his first few observations went well, and the teachers were given positive results. Those first four or five observations that he conducted all were scored in the upper 2’s, all proficient ratings.

Although the first several observations were rated as proficient, Mr. Boudreaux’s worst COMPASS evaluation also occurred during the 2012-2013 school year, the first year of using the current COMPASS system. It was an announced observation, but the teacher “still failed to get the pieces together.” Although the teacher received an ineffective rating on that lesson, she was able to “squeak out” a one point five for a final rating at the end of the year, the lowest effective rating one can receive on the COMPASS rubric. “Whenever the lesson plan is lackluster and the student motivation is just not there and at any given time there are nine or ten kids that are just completely disengaged, the wheels come off the train, and it’s just, it’s terrible. You can sort of feel the pressure mounting on the teacher.” Mr. Boudreaux recalls the post conference as an exhausting one in which the teacher disagreed with him on some of the ratings. “She was rather defensive about the lesson plan

because it looked a lot like other people's lesson plans and I don't think to this day she understands that the, what's written on paper is the lesson plan but what you have in your mind that you're going to do is sometimes the difference between an effective thing on paper and an ineffective thing on paper." Sometimes teachers can compose thoughtful lesson plans but have issues with executing these plans, especially if the plans do not consider the students' needs or the teacher's teaching style. This teacher was upset during the post conference, acted defensive towards her scores, and demonstrated similar issues during walkthrough observations. Overall, Mr. Boudreaux knew that if this teacher could not teach an effective lesson even when she knew he would be observing her, she probably struggled on a daily basis to provide effective classroom instruction. Mr. Boudreaux works diligently with his teachers to improve their practice so for a teacher to be defensive and unwilling to receive feedback from an observation frustrates him. Her defensiveness was heightened by the knowledge that her scores were tied to her tenure and pay, creating a high-stakes situation.

Even though it was an ineffective lesson, at least Mr. Boudreaux was able to observe what was presumably occurring on a daily basis in that classroom. In opposition to this issue, it can also be during announced observations that evaluators see teachers who know the rubric and plan their lesson to fit the rubric for that particular lesson. According to Mr. Boudreaux, "it's obvious we've been doing this for two years and the teachers, during announced observations, know exactly what the verbiage on proficient and highly effective look like and they tailor lessons to hit all of those things." Teachers are now aware of how to plan lessons that adhere to COMPASS standards and they realize their jobs could be at stake, so some are manipulating their lessons to purposely score well on COMPASS. This is

especially true for their one announced observation, when they know about the observation prior to that date. This can be frustrating to administrators like Mr. Boudreaux who try to use their role as evaluator to also coach teachers on their practice. It can be difficult to elicit true data on a teachers' weakness if the teachers are preparing lessons specifically modified to score well on COMPASS and not in their own true teaching style that is used at other times. One thing that helps to not inflate teacher ratings and give a more accurate depiction of his classrooms is the use of walkthrough observation data. "Luckily LSD has given me...the directive to use all data that we garner from a teacher's walkthroughs to come up with the final evaluation." This means Mr. Boudreaux can examine any walkthrough data on the teacher from that year and use this evidence in combination with formal observation data. This is Mr. Boudreaux's response to teachers purposely writing plans to score well on COMPASS for just one lesson, on the day of their announced observation. By using walkthrough data, he is able to ensure a more accurate depiction of that teacher's effectiveness is captured.

Mr. Boudreaux views his role as a COMPASS evaluator to give teachers "the feedback they need to become better teachers" by "identifying strengths and weaknesses" which is what he sees as the purpose of COMPASS. He has learned to creatively use COMPASS, which does not mandate a high degree of coaching within the policy, to carve out a more coach-oriented role as evaluator. Conducting walkthroughs and formal observations that are scored on the COMPASS rubric "does not work well in developing a relationship of trust with staff members...it has with it a bit of a punitive side that I can't avoid." Although he tries to use COMPASS observations as sessions to improve teachers' effectiveness in the classroom, teachers still know that their reputations, pay, and job security are all on the

line. So in addition to unannounced COMPASS walkthroughs and announced formal observations, he has begun a practice of informal walkthrough observations that have no connection with their scores on COMPASS. He uses his iPad to score COMPASS walkthrough and formal observations. When a teacher sees his iPad out, they know they are being formally evaluated, either in a walkthrough or full evaluation. When he does not have his iPad with him, the teachers know that he is there informally to see what they are doing, “and it’s not part of their COMPASS observation.” They know they will have a verbal conversation with him later about their practice. He describes his roles as evaluator and coach as “a very, very difficult line to walk” but “definitely a practice I’m going to continue.” He plans to continue his coaching of teachers through informal, non-punitive, non-COMPASS walkthroughs next year and “to have more conversations about what we need to improve and how we need to get better.” Mr. Boudreaux uses COMPASS as a tool to further his teachers’ effectiveness, but he recognizes its “punitive” nature as one that can cause teachers to become defensive and less open to his coaching. He combats this by extending his types of observations outside of the realm of what is required by COMPASS in an effort to reach his teachers. Mr. Boudreaux can clearly describe and define his roles as a high-stakes evaluator. He is, at heart, an instructional coach who desires teachers’ trust of his expertise and an evaluator who uses high-stakes evaluation to drive the focus of his school.

**Mr. Breaux’s Story: Striving for a fair and accurate way to evaluate teaching effectiveness**

Mr. Breaux is principal of the highest performing middle school in LSD. This is his second year as the principal. The school received an A on their school report card for the 2013 school year, and forty percent of its students are on free or reduced lunch. The school

is located in a remote community and has less than three hundred students. Mr. Breaux has one assistant principal who assists him with COMPASS evaluations. He recalls his own first teacher evaluation as a first year teacher when he scored a 100% on the “old rubric” twenty years ago. He questions why everyone automatically received that rating, even first year teachers, because “really? I wasn’t that good.” He knew something needed to change because teachers were in need of “real, actual feedback” from their evaluations and also because “we’ve got some teachers [in education] that we need to get rid of.” Mr. Breaux is a former athletic coach who wants to follow the rules and be provided with clear guidelines about the implementation of COMPASS so that he is as fair and accurate as possible, and he is frustrated at the information he has been given about COMPASS from his initial training through year two of implementation.

As a former athletic coach, Mr. Breaux wants to follow the rules and make sure other evaluators are doing the same. Mr. Breaux is bothered that no time has been spent on ensuring inter-rater reliability across the district and is concerned that “everybody in this parish...had a different interpretation of what a one, two, three, and a four were after that training.” This is a concern to him because he wants his implementation of COMPASS to be fair and accurate. He points to evidence of a lack of inter-rater reliability when he discusses 1c on the COMPASS rubric “Setting Instructional Outcomes.” Mr. Breaux is troubled that administrators in the district are scoring this indicator differently because some rate this component of the rubric prior even to observing the lesson, based on the lesson plan, and others, like himself, are more concerned with the “outcome” of the lesson or “what is produced by the kid in that lesson.” When initially learning COMPASS, Mr. Breaux communicated frequently with his assistant principal so they could “define what those

attributes were going to mean” for their school as they observed, scripted, and scored many of those first observations together; he does recall it to be a very time-consuming process. Part of the learning process was trying to navigate the rubric, which he describes as “not linear” because “it’s not a continuum for some of these attributes”, and this leads him to question “where do we put this”? This is similar to the wait time example given in Mr. Boudreaux’s story when he described the rubric as lacking alignment. Some attributes only appear in one level, causing principals to wonder how to score it if it should appear in another level. He also describes some of the “contrived” lessons in Physical Education, Adapted Physical Education, and Community Based Training as “fake” because COMPASS “just doesn’t really fit.”

Last year Mr. Breaux focused on becoming more comfortable with the COMPASS instrument and as he became more knowledgeable about using the rubric, he found what he considers a flaw in the rubric: it sets the bar for a passing score too low. He saw a lesson that was disappointing, and he, in turn, found the teacher difficult to coach because of her defensiveness. The students “were not interested in cooperating” which is rare at his school because the students are generally compliant. The students looked at the teacher “as if she was crazy” when she asked questions and expected them to answer. Mr. Breaux assumed from this that the teacher was only asking questions during the lesson because he was there to evaluate her. During the post conference, the teacher was extremely defensive about her practice and disagreed with her score. Mr. Breaux found it difficult to reach her because “her perception of reality was way far off from what the actual reality was.” The teacher passed the observation with a 1.8 rating despite the observation being “like pulling teeth.” Although Mr. Breaux scored the teacher as accurately as he could, he did not agree

with the overall rating, which was a passing score. He argued that “if you stay awake during your observation, you’re not going to fail,” which is one fault that he finds with the COMPASS rubric. Because of the way COMPASS is scored, it is extremely difficult for teachers to receive an overall Ineffective rating. It is demanding for him, as a principal who self-identifies as an instructional coach, to use COMPASS as a tool to improve teaching when the overall score gives ineffective teachers a false rating and helps them remain as teachers.

Even though it is tough to terminate ineffective teachers using COMPASS and he struggles in using the rubric for some teachers, it allows him to push his teachers by setting high expectations and asking his teachers “What are some of the things you can do to try to attain these [higher] levels?” Although he finds the rubric sets a passing score too low, he still feels that it has improved teaching practices and pushes teachers to improve their practice in ways that the first rubric he was evaluated on as a teacher never did. Mr. Breaux and his assistant principal conduct walkthrough observations frequently using one component of the COMPASS rubric at a time, and he provides feedback through email and in conversations. To him, COMPASS is a major part of being a principal in Louisiana today, and he uses COMPASS as a coaching tool, which is what he views as his main responsibility as a principal. His past coaching experience helps him to “encourage their strengths...and...fix their weaknesses.” As a former athletic coach who wants to follow the rules of the playbook, he is frustrated by a lack of guidance and answers about how to use COMPASS fairly and accurately. He uses the critical attributes of Highly Effective to guide his coaching and he feels satisfied when his best teachers score a Highly Effective rating. He strives to implement the COMPASS model with fidelity, although he becomes frustrated

when issues with the model occur and when it is difficult to use the rubric in some situations.

### **Ms. Robichaux's Story: Begrudgingly following an unhelpful policy**

Ms. Robichaux is the principal of the highest performing high school in LSD with a school report card grade of B for the 2013 school year. Nearly fifty percent of its 1200 students receive free or reduced lunch. This is her second year as the principal, and she is one of five administrators at her school who use COMPASS. In her previous position, she used the old district rubric to evaluate teachers, and she remembers receiving extensive training in three to five days and passing a test of reliability before being able to use it. Through the use of this instrument, she “got rid of a tenured teacher with 28 years of experience.” Ms. Robichaux liked the old rubric and feels like it served its purpose in allowing administrators to focus only on struggling teachers who really needed to be evaluated annually, as compared to COMPASS which forces all principals to evaluate their entire faculty twice a year. It “gave teachers an idea for improvement” because it was “so cut and dry.” She was comfortable with this method of evaluation and used it as a tool to increase teacher effectiveness. In stark contrast, she feels that she and her team are merely “going through the motions” of COMPASS as part of a mandate that does not fit with her vision of improving teaching practices at her school.

At the beginning of the second year of COMPASS implementation, Ms. Robichaux's teachers participated in Professional Learning Communities in which they “took apart the COMPASS rubric piece by piece,” so they now understand “it's not really what the teacher is doing, it's what the students are doing.” The English teachers at her school now realize that using the teaching method of a Socratic Seminar will score them a highly effective on the



COMPASS rubric because it is “student directed” and “all higher order.” Because of this new understanding, “the scores have gotten better” during year two in comparison with year one. This makes it unclear whether actual teaching practices have improved or if teachers are simply learning the best methods to achieving high scores on COMPASS and adapting their lessons to fit the rubric. These contrived lessons for announced observations can make it difficult to actually improve teaching practice.

During year one of implementation, the teachers did not understand or know the COMPASS rubric enough to adjust their lessons to fit the rubric. This caused the school’s evaluators to rate very few teachers as Highly Effective. A challenge came at the beginning of year two when the local newspaper printed a report from the Louisiana Department of Education that “touted [her school] as one of the schools that had very few highly effective teachers.” The Louisiana Department of Education “said it was a true picture of how the COMPASS rubric should look” because they were an A school with a small number of highly effective teachers. Ms. Robichaux “had to explain that one” to her teachers, who were upset that they were perhaps scored too harshly. Ms. Robichaux and her team were following a district directive that a teacher has “to have all proficient and a majority of the highly effective ratings to be considered highly effective” which is a conflicting instruction with those “from the state” which was to rate a teacher by “whichever target level had the most checks.” It is up to principals to interpret much of the implementation of COMPASS. Some principals may be comfortable with this flexibility to interpret the rubric in a way that fits with their school, but others like Mr. Breaux and Ms. Robichaux want clear directives because they are concerned with evaluating their teachers fairly and accurately because of its relation to teacher pay and tenure.

Even at the end of year two of COMPASS, there were still inconsistencies. At an administrative meeting, one of Ms. Robichaux's assistant principals mentioned that she had evaluated two Highly Effective teachers this semester, one with a 3.48 that she rounded up to a 3.5. All five administrators disagreed about whether or not a score could be rounded. Ms. Robichaux placed a call to a district supervisor, who said yes, a score can be rounded up. She then looked over her COMPASS training manual that she received from the state and found no, a score cannot be rounded. The assistant principal questioned, "How many other schools are rounding?" Ms. Robichaux is unsettled by the inconsistencies in COMPASS, largely because "teacher pay is associated" with it. Again, Ms. Robichaux discovered an issue with the implementation of COMPASS that was lost in translation from state to district to school level administrators.

Ms. Robichaux is passionate in her negative feelings about COMPASS. She believes it has "given free reign to ineffective teachers to stay in the teaching profession." Because of litigation surrounding COMPASS, she "could not get rid of a tenured teacher right now" who she "wrote up three times this year" and who "failed her observations." She recognizes her duty as an administrator to be an evaluator and feels like her greatest responsibility as principal is to be "an instructional leader" but COMPASS is a constricting evaluation system for a principal who wants more autonomy to evaluate only teachers who are struggling annually because her administrative team doesn't "have time to help the teachers who really need the help." Using COMPASS the past two years has been time-consuming and they are "going through the motions" of COMPASS in an effort to "just...get things done." Her new role as a COMPASS evaluator has forced her to abandon some of her previous practices of working intensely with struggling teachers because of the sake of time. To her,

COMPASS has impeded her role as an instructional coach, and she implements COMPASS only because it is part of her obligation as a principal. COMPASS evaluator is a role she is forced to perform despite her conflict with the policy itself and the lack of clarity in its implementation. She's frustrated by the inconsistencies she has discovered in the implementation of COMPASS and finds it an imposition that is disjointed with her own vision of her leadership role.

### **Mr. Richard's Story: Following the policy because it is required**

Mr. Richard is the principal of a middle school with approximately 500 students, 50% of whom are recipients of free or reduced lunch. The school received a B letter grade from the state of Louisiana in 2013, and Mr. Richard has been the principal for five years, although he has over thirty years of experience in education altogether. Mr. Richard has experienced four different teacher evaluation systems throughout his career, so his mindset has become to "go with the flow" confident that he'll "figure out how it works" for him and his school.

In adapting COMPASS to fit his role as an evaluator, Mr. Richard decided not to continue scripting every part of a lesson as he was initially trained to do. He found it cumbersome to "go back and look at every line" that he observed during a lesson, so he began selectively scripting only the evidence he knew was pertinent to the COMPASS rubric. This way, when evaluating questioning, he had all the questions asked during the lesson in one area of his script. He finds scripting every word of a lesson unnecessary to the process of formal observations. This has improved his efficiency in conducting formal observations, which is really important to him because he was short-staffed in administration during the second year of COMPASS implementation.

The worst COMPASS observation he ever conducted was a teacher who did not teach the lesson she had submitted in her plans for the day. Mr. Richard always looks at a teacher's plans and decides on the best day to observe them based on those in an effort to fairly see each teacher at his or her best. Instead of teaching the lesson planned, this teacher had the students "catch up on their journal topics that they had not completed." He questioned himself about how to handle the situation: "Do you stay in the class and continue the observation? Do you say hmm, because she's going to fail it?" Mr. Richard decided it was fair to stay and use the lesson as a COMPASS observation, and the teacher did fail the observation. "And so that's tough, but it's not tough because it's part of COMPASS. It's tough because she just wasn't being effective at that particular time." The teacher could have taught the lesson after seeing Mr. Richard walk in, but chose to continue with allowing the students to complete their overdue journals. She later told Mr. Richard that the nine weeks were about to end, and because the students never complete their assignments when they need to, she had to give them the time to finish so she could submit the grades. Mr. Richard felt fully justified in scoring the lesson as it was on the COMPASS rubric because "she obviously wasn't prepared to teach that day." Mr. Richard does not let teachers know when their formal unannounced observation is, and when he enters, they are not aware of whether it is their formal observation or an informal walkthrough. They only realize if he stays for more than the typical 5-10 minute walkthrough observation time. In this case, the teacher later told Mr. Richard she did not know it was her formal observation and did not realize Mr. Richard would be staying in the room the entire lesson. In general, Mr. Richard does not conduct face-to-face post conferences with the teachers he observes because emailing the scored COMPASS rubric is "a time-saving tool" because he is

not “spending an hour for a teacher to explain why it is that she didn’t score well.” The most difficult lessons to score are “the ones that you really have to figure out how to be professional and not, you know, put anybody down.” This occurs when there is simply a complete lack of evidence for a component of COMPASS.

It’s when you have to figure out how to nicely tell somebody that you know, you just don’t have it. And you just didn’t get it today.

It wasn’t there...It’s when I come back from a classroom and I look at my notesheet and there’s nothing. I have no evidence. There are no questions. There were no questions asked...How do you be nice in telling them they’re ineffective in asking them questions?

Mr. Richard is concerned about staying on good terms with a teacher even when he has to criticize their practice through COMPASS evaluations. He strives to be nice to teachers even when their practices are ineffective. When any component of a lesson is ineffective, Mr. Richard always provides “a comment on what to do to make it proficient.” He makes it clear through his email about the teachers’ scores that they are welcome to see him with questions. He does note that the majority of teachers do not ask to see him face to face or question their scores.

Mr. Richard sees his role in COMPASS “to ensure that all of the components are there” and that all of the deadlines are met. By components, he is referring to all of the documents that must be provided by teachers and principals through the COMPASS process. This includes a teachers’ Professional Growth Plans and Student Learning Targets, written each year by the teacher. He follows COMPASS policy and completes what is required of himself and his administrative team, although he doesn’t have any strong

feelings for or against the teacher evaluation system. He does not go beyond what is formally mandated by COMPASS because of the time consuming nature of teacher evaluation.

### **Ms. Simoneaux's Story: Supplementing COMPASS with Coaching Knowledge**

Ms. Simoneaux is an assistant principal at a middle school with approximately 400 students with a 64% free or reduced lunch rate. The school was rated a "B" school during the 2012-1013 school year. This is her first year as the assistant principal. Before this, she served as a TAP Master teacher at another middle school that uses the TAP teacher evaluation system instead of COMPASS. She is open in sharing her feelings about COMPASS in relation to TAP from her first TAP training to her use of COMPASS currently. The 2013-2014 school year, the second year of COMPASS implementation, was her first year using COMPASS.

The first TAP training she received from the Louisiana Department of Education was eight full days, and at the end she was required to pass a test certifying that she could score a lesson using the TAP rubric accurately. Following this training "when I got into doing an actual observation at my school I just felt like I knew what to do because I had passed this test, you know, and I was trained." In comparison, her COMPASS training was "2 to 3 days over the summer" and "a weak version of the TAP training." By the end, she "just didn't feel ...comfortable enough in the actual COMPASS rubric." She can tell that the COMPASS rubric is only pieces of the Danielson Framework, as the current rubric feels incomplete to her. At times she is not sure where evidence fits or whether a person is proficient or emerging. From the beginning, she has used her TAP training to help fill the holes in COMPASS. In addition to scoring a teacher on the COMPASS rubric, Ms. Simoneaux used the post

conference procedure from TAP to give her teachers a strength and a weakness. A post conference with Ms. Simoneaux can be time-consuming as she highlights a strength and determines a “weakness...and give them how that lesson could have looked different.” The power of a post conference is one of “individualized professional development” that Ms. Simoneaux bases as one of her primary responsibilities as an assistant principal.

The power of the post conference was evident during Ms. Simoneaux’s very first COMPASS observation. She observed a first year teacher at the beginning of the school year who was lacking the concept of backward design, but had a lot of potential.

So I sat down with her and I actually went through, you know, how this lesson could have looked a little different if you would have just done this, and she went from emerging the first go round to this go round when I observed her again in the spring, she went up to Proficient. So she grew.

Ms. Simoneaux was proud that she was able to impact this teacher’s scores through the post conference and walkthrough feedback. To her, this is her most important responsibility as a principal and is her role as a COMPASS evaluator. This is a unique perspective in comparison to the other participants who either do not conduct post conferences (Ms. Richard) or did not mention them as an important part of the teacher evaluation process.

Ms. Simoneaux strives to spend extensive time in meeting with teachers during post conferences and providing individualized feedback, but time does become an issue. “The rest of the school still has to run even though you have a post conference.” Her role as a strong instructional coach does become time consuming. Her previous role as a master

teacher allowed her to prioritize professional development, whereas her role as principal is split between much more than teacher evaluation activities. The end of the school year is a particularly busy time, and it has been difficult to fit in the final COMPASS evaluations, but she strives to complete a full post conference for all of her observations. In the next year of COMPASS implementation, she plans on continuing this practice but also working on tying walkthrough observations into observations. She also would like to leverage her “TAP knowledge” in more in an effort “to help take COMPASS...and make it a little better.” Her conviction about post conferencing as a vital tool to impact teacher performance demonstrates that teacher evaluation is important to her; however, she does feel that the COMPASS rubric is simply “bits and pieces” of what should be the bigger picture, or the Danielson Framework in its entirety. It is clear that her previous TAP experience helps to guide her as a coach in her current role as a COMPASS evaluator because the TAP teacher evaluation model places heavy emphasis on mentoring and coaching to improve teacher evaluation scores.

**Mr. Thibodeaux’s Story: Evaluating his teachers to evaluate himself as leader**

Mr. Thibodeaux is the principal of a “B” high school in LSD with approximately 1300 students, around 50% of whom are considered free or reduced lunch. This is his second year as principal; he previously served for two years as the principal of a middle school in LSD. Mr. Thibodeaux is a soft-spoken man who is obviously thoughtful and conscientious about the statements he makes regarding teacher evaluation. His background is as an elective teacher, and the majority of his COMPASS evaluations are of elective teachers like Agriculture, Physical Education, and Business.



Mr. Thibodeaux wants COMPASS to be implemented fairly at his school. He has worked with his administrative team to practice scripting and scoring video lessons. He emphasized working together to his administrative staff in order to be “as much on the same page as possible” and avoid teachers wanting “to get evaluated by this person because they look at it different or expect different things.” He wanted consistency among his staff. For the teachers, he provided professional development on the rubric itself to make them more comfortable with the expectations.

His first few observations took a while to complete as he learned the rubric and made “sure I was doing it the right way.” He “still had a lot of questions going into it” but realized that “the teachers were still getting acclimated with the rubric” as well. The first few lessons scored in the Effective: Emerging range, but he has seen a “change in mindset” as teachers are “letting go a little bit more within the classroom and allowing kids to take the lead on some areas.” This was a difficult concept for Mr. Thibodeaux because when he was a teacher he liked control and didn’t like a whole lot of talking in his class. He also has “never been a teacher observed on the COMPASS model,” so he has learned COMPASS from the evaluator’s side.

The walkthrough procedure was also changed to meet the expectations of COMPASS. Each teacher at the school is observed informally weekly by a different administrator and given feedback. It’s crucial to Mr. Thibodeaux to always give each teacher some positive feedback and a statement of what they could work on. These walkthrough evaluations are on one component of COMPASS each week, and they are used as evidence during formal evaluations. This increases the reliability of the COMPASS evaluation. If a teacher “knows we’re coming...and that one day, she does a good job with

questioning but the other thirty times we've been in there she didn't do it because she didn't know we were coming...we use that as evidence." Between walkthrough and formal observations, "it's hard to be ineffective in COMPASS," so Mr. Thibodeaux has not had a teacher fail COMPASS. He has seen a disaster of a lesson in which students were "off task" and the teacher had "no control." Students never completed "any independent work" nor were they assessed. Despite his dissatisfaction with the lesson, it still did not fall in the ineffective category when he rated it as accurately as he could. He has worked with this teacher from the first year of COMPASS, and has seen some areas of her practice improve during the second year of COMPASS implementation.

Mr. Thibodeaux sees himself as a resource for his teachers that he observes, and he knows that being compassionate about the learning curve is crucial for impacting his teachers' performance on COMPASS. He tries to understand that COMPASS is new to teachers just as it is new to the administrators at his school. To him the purpose of COMPASS is to increase student engagement and make students more responsible for their own learning, with teachers as the facilitators. He holds himself responsible for his teachers' success on COMPASS because "if a teacher is not doing what they're supposed to do in the classroom, it is my job to get them where they need to be." His evaluation of teachers then becomes an evaluation of himself and how well he communicated his expectations to his teachers. A "good teacher...will do well on any instrument" that is used to evaluate them. He accepts COMPASS as a tool that evaluates his teachers' effectiveness and his own leadership.

### **Ms. Fusilier's Story: Navigating outside influences on her role as evaluator**

Ms. Fusilier is one of Mr. Thibodeaux's assistant principals, and this is her second year at the high school level. Her background is as a teaching coach and master teacher on the middle school level. She oversees and evaluates the English department. She sees her role as an administrator as someone who has to make "those hard, tough decisions sometimes", but she accepts it as a part of the job. Some of those tough decisions occur during COMPASS evaluations when she has to score teachers on their lessons.

She was introduced to COMPASS during her second week as an administrator. Her first observations using the COMPASS rubric "took awhile to get through" because she didn't understand "what some of the components of the rubric really meant." In time, Ms. Fusilier began to feel more comfortable with using the rubric, although she just felt that it "left a lot to be interpreted." Ms. Fusilier does her best to interpret the rubric as fairly as she can. She often compares student test scores to her COMPASS scores because it confirms her thinking. This is usually long after the teachers receive their COMPASS evaluation scores, when student test data is received at the end of the school year, but she feels that her observation scores are validated if there is a correlation between those scores and student achievement data on high stakes tests. According to the basic premise of teacher evaluation, this should be the case; however, Ms. Fusilier has discovered that to not always be true in her evaluations. In some instances teachers with low evaluation scores have students score well on their high-stakes tests.

During the first year of COMPASS, Ms. Fusilier did not score any observations as Highly Effective, but during the second year, three of her teachers received Highly Effective scores as she has "definitely seen some growth" within her department. Her most impressive evaluation was an unannounced one in which she could see evidence that the

teacher had specifically worked in the areas she was given as refinement areas during the first year of COMPASS. Ms. Fusilier was impressed by the level of engagement and the student-centered lesson in which it was obvious that the students would be prepared for college. One component of the COMPASS rubric was not apparent in this lesson: students asking each other higher order thinking questions; however, Ms. Fusilier was able to use evidence from “ten documentations...of students asking questions of one another” as supplemental sources of COMPASS data. She likes that COMPASS allows her to give credit to teachers for areas that may not be visible in the one formal lesson observed. This demonstrates that some of the interpretation of COMPASS allows principals the flexibility to score teachers in the way that they deem is most fair and accurate.

Another English teacher had many walkthrough observations that were not that impressive. Ms. Fusilier gave him extensive feedback throughout the year, but he submitted a lesson plan from the previous teacher and obviously did not follow it. “He was winging it. He was all over the place.” Ms. Fusilier was disappointed because of the time and effort she put forth in coaching this teacher. She didn’t realize until his formal observation that he obviously was submitting another teacher’s lesson plans without revising them or following them at all, so “he went wrong in planning” which led to a scattered lesson in which the kids were bored. However, this teacher did pass his observation because, as evidenced in the previous stories, it’s very difficult to be ineffective in COMPASS. At the same time, Ms. Fusilier treaded lightly through his post conference because she doesn’t want him to resign from his teaching position. Even though she gave him many areas to work on, she made sure to give him positive comments as well, like that “the students were well-behaved.” She feels like he’s worth saving as a teacher, and she doesn’t want him to

become discouraged. He was alternately certified, but he still has a lot to learn. He was polite and accepting during the post conference, but Ms. Fusilier is “watching now with eyes wide open” to see how he performs in the future. Ms. Fusilier was hoping his students’ standardized test scores do not reflect what she saw in his teaching effectiveness. Because teacher shortages are not uncommon in LSD, Ms. Fusilier knows that even if this teacher has significant weaknesses in his practice, if she feels she can coach him to become a better teacher, it is worth her time and efforts.

One of her teachers has seen great growth in student test scores, but he did not score in the Highly Effective range on either observation. He is not “a fancy flashy teacher” who tries to impress Ms. Fusilier. She sees evidence of great rapport with students and knowledge of his test and content area. His lessons are nothing fancy and when walking through informally she doesn’t feel like “oh my god, this was so good,” yet the students are always working diligently on a meaningful task. Because he does not use the best or newest strategies he does not score in the Highly Effective range, but his students brag about increases in their test scores and attribute their improvement to their teacher. Ms. Fusilier learned from this experience because she has learned that what he does works for his students and he is an effective teacher even if he is not Highly Effective according to the COMPASS rubric.

At the end of this school year, Ms. Fusilier was contacted by another teacher’s students and several parents who complained that their teacher wasn’t “preparing” them. Following this, Ms. Fusilier had to conduct her final COMPASS observation of this teacher and she found it difficult to separate her conversations with the students and parents with the evidence from her observation. During the first year of COMPASS, this teacher struggled

with management, and Ms. Fusilier also questioned herself “because I was just so impressed that he had things under control, was I neglecting to see other things?” She went back to the 24 walkthrough observations she had conducted on various lessons throughout the year and saw evidence of his effective teaching. However, she spotted a trend that most comments were about student discussion and questioning. The student complaints were that all they did in class was “discuss and talk” without “getting anywhere.” Ms. Fusilier scored his lesson as proficient based on her evidence from the formal observation and walkthrough observations. During her post conference she brought up the trend she saw in the data from observations. She decided she could not “let those outside influences change what the data said.” In this case, his high student test scores helped to confirm what she saw during his observations. If evaluators are not careful, it is easy to allow outside opinions influence ratings.

Ms. Fusilier has learned a lot in her first two years as an administrator through the implementation of COMPASS. She thinks of evaluation as a way to hold teachers accountable but also to “measure their growth” from year to year which is something that she has seen evidence of during her two years in using COMPASS within the English department at her school. She uses COMPASS observations and student performance on standardized tests to “decide who’s going to teach what.” Ms. Fusilier looks at the big picture of evaluation, and although she has tried to see correlations between evidence she views in the classroom and student performance scores, she has found several cases of inconsistency. Although this has caused her to doubt her own proficiency in her role as evaluator, she can cite large amounts of her own observation data to corroborate with her observation scores. She is a reflective evaluator and coach who uses observation data in

her decision-making but doubts the COMPASS process after her past two years of experiences.

### **Thematic Analysis: The Roles of Principals Implementing COMPASS**

The narratives above represent the stories of the seven participants as they described their roles in COMPASS teacher evaluation. Within their stories, their part in COMPASS can be divided into two main functions: *Instructional Coach* and *High Stakes Evaluator*. These principals feel the pressure of conducting high stakes teacher evaluations that impact teacher pay and tenure. Ms. Fusilier mentioned hearing that other principals were “uncomfortable” with “making a judgment call on somebody’s pay.” It has not been the norm for principals to wield this type of power, and most educators, unless they have worked outside of the field of education, are unused to this type of role. However, most participants seem more comfortable with the idea of coaching their teachers or using the evaluation as a coaching tool than just evaluating their teachers.

*A priori* coding was initially used to analyze the data based on Fiedler’s (1973) contingencies, but several more codes emerged during the analysis. These codes were then classified into one of the three contingencies: *Task Structure*, *Leader Member Relations*, and *Position Power*. From here, *Position Power* was renamed as the role *High Stakes Evaluator* because the role of evaluator yields with it power over teacher pay and tenure through the nature of the role. *Task Structure/Leader Member Relations* was renamed as the role *Instructional Coach* because within it principals used their role to instruct teachers about the tasks they were performing in the classroom and at the same time navigated their relationships with their teachers, or members of their organization. The overarching theme is one of the dimensions of Hallinger and Murphy’s (1985) Instructional Leadership

Theory: *Managing the Instructional Program*, and all three contingencies can fall under this umbrella as principals navigate their primary roles in COMPASS as *High Stakes Evaluator* and *Instructional Coach*. They are managing the instructional programs at their school sites by evaluating and coaching teachers. This framework helped me in answering my research question by clarifying the codes that separated principals into their two main roles of *High Stakes Evaluator* and *Instructional Coach*. In the following sections, each role will be defined by the words of the participants as they describe their experiences as COMPASS evaluators.

### ***Instructional Coach***

The role of *Instructional Coach* was made up of two dimensions of Fiedler's Contingency Theory (1973): *Task Structure* and *Leader Member Relations*. Task Structure includes directing the employees in how to perform a task, so this would include any time the principals instructed teachers on the COMPASS rubric and how to perform well on each component. Leader Member Relations refers to instances when when the principals supported teachers through encouraging words or positive feedback. It became obvious during data collection that Instructional Coach is a role that the most participants self-identify as a primary obligation of principalship and directly tied to COMPASS teacher evaluations.

The participants mentioned many instances of coaching and supporting their teachers through the COMPASS process, both formally and informally. Informal verbal feedback is given to teachers as well as more formalized post conferences based on the teacher's formal observation. The participants worked to set clear expectations for their teachers using COMPASS and provide professional development for their teachers on the new COMPASS rubric and process. It is evident that the principals care about their teachers



and their performance on COMPASS. They want the teachers to do well and to improve on their observations.

One component of being an *Instructional Coach* is to conduct walkthrough observations and providing teachers with feedback, something that came up many times during each interview (Ovando & Ramirez, 2006). Because the participants identified walkthroughs as a significant task associated with COMPASS and their role as Instructional Coach, it deserves a closer inspection as to how each participant conducts walkthrough observations. Although all the participants discussed conducting walkthroughs, they all used different types of measurement and utilized the data from walkthroughs for different purposes. The information from each participant about their use of walkthroughs is located in Table 2.

**Table 2: Use of Walkthrough Observations**

<b>Participant</b>	<b>Instrument</b>	<b>Use of Data</b>
Mr. Boudreaux	All 5 components of COMPASS rubric	LSD “has given directive to use all data we garner from a teacher’s walkthrough” in combination with formal observation
Mr. Breaux	1 component of COMPASS at a time on a weekly rotation	Used to “break a tie” and “as an average of all data” for a formal observation
Mr. Robichaux	All 5 components of COMPASS rubric	Used with formal observation
Mr. Richard	Examined different areas: management, lesson plan check, questioning without the use of a rubric	“kept separate” from formal observation
Ms. Simoneaux	All 5 components of COMPASS rubric	“Used somewhat with COMPASS” formal observations and “working to use with more fidelity”
Mr. Thibodeaux*	1 component of COMPASS at a time on a weekly rotation	Used with formal observation
Ms. Fusilier*		

\*same school

As evidenced above, five participants always use walkthrough data in combination with formal observation data, one participant is working to use this more, and one participant keeps walkthrough observations completely separate. Three participants use one component of COMPASS at a time, three use all five components of COMPASS, and one participant uses various topics of COMPASS without the COMPASS rubric. This demonstrates the different interpretations of policy by principals in how they conduct walkthrough observations and the use of the data. According to Mr. Boudreaux, the LSD “has given directive to use all data we garner from a teacher’s walkthrough.” For example, if students did not ask each other higher order questions during a formal observation, but Mr. Boudreaux observed this in several walkthrough visits in prior weeks, he can use that

evidence to give the teacher a higher score. Mr. Boudreaux calls it a directive from LSD, but the data from other principals shows that they feel flexibility with the directive. All of the participants discussed walkthrough observations as an important part of their instructional leadership, but they are gathering and utilizing the data in different ways from school to school.

**Using informal conversations with teachers to coach.** All of the participants also spoke about using walkthrough observations to give teachers feedback, providing professional development, and conversing with their teachers about improvement of their practices. Mr. Boudreaux clearly defined his role as *Instructional Coach* by shaking his head no or walking in with his hands in his pockets when he entered a room to signify to teachers that he was there as a coach and not an evaluator. He does this because “teachers don’t function well in an environment when they’re always being formally evaluated. Sometimes they just want to have a conversation.” He tries to verbally give feedback to his teachers often, but “the difficult part is communicating that feedback in a way that is meaningful.” He wants teachers to be able to take his feedback and actually apply it in their practice, rather than just agree with his comments.

Both Mr. Boudreaux and Mr. Breaux had experiences of teachers who performed poorly on a formal observation blaming the observer for their results. Mr. Breaux described his teacher as being out of touch with reality, so she did not respond well to his coaching because of her defensiveness. Mr. Breaux focuses on coaching his teachers daily and provides “support to help them from where they are...and bump them up a level or two”, no matter how poor their performance is. He tries to set clear expectations by “working with them, teaching them, being able to clearly define and explain what it is that we want from

them.” Mr. Breaux coaches his teachers on a daily basis at his small middle school, but at a larger school that could be more challenging.

**Using their backgrounds as teacher coaches.** In stark contrast, Ms. Simoneaux finds a face-to-face post conference with teachers a powerful tool for personalized professional development. She uses her background in TAP to inform her post conference style by giving each teacher an area of strength and an area of weakness, as well as a specific way to correct the weakness (the TAP protocol for post conferences). She uses walkthrough observations as evidence at times, and she discusses walkthrough feedback more in depth during her post conferences with teachers. As a former TAP master teacher, she uses her experiences of coaching teachers in her work with COMPASS as an assistant principal. Because of her other responsibilities as an assistant principal, to her the post conference is the main avenue of fulfilling her role of *Instructional Coach*.

Ms. Fusilier, like Ms. Simoneaux, comes from a background in teacher coaching, and she has continued that in her role as assistant principal, so being an *Instructional Coach* has felt natural to her. She always tries to encourage teachers by sharing what they “have going for them” as well as what they have to work on. She takes her job of mentoring teachers seriously, and she conducts a large amount of walkthrough observations on her teachers to provide them with feedback and use as evidence for formal observations, as demonstrated in her story about the teacher she observed who had student complaints. She closely monitors the group of teachers that are assigned to her and works diligently with them to improve their practice. She takes her work with teachers seriously and found herself disappointed in the teacher who did not show improvement after she worked with him closely. She looks to find correlations between what teachers score on COMPASS

observations and how students perform on high stakes exams to provide evidence if she is fulfilling her role as coach and evaluator. She doubts her own competency in her role as evaluator when the two do not align.

**Supporting teachers through coaching.** Mr. Thibodeaux was the only participant to discuss how he correlates his teachers' success on evaluation with his own value as a leader. As a former elective teacher, he mostly observes the elective classes at his high school. In working with the teachers he observes formally he tries to "use tools in order for them to feel more comfortable with the instrument" and "give them suggestions on how to do things." Setting clear expectations is very important to him since he sees his teachers' evaluations as a direct reflection of his own skill as a coach. He values communication with his teachers, and sees any failure among his teachers' performance as a personal reflection upon himself as a principal, so he works to provide the resources and support that those teachers need. He sees his role as a high stakes evaluator in correlation to how his teachers score on COMPASS. His teachers' performance is impacted by his interactions with them through his coaching and he uses their performance ratings as a sign that he is fulfilling the obligations of his own self-defined role as evaluator.

**Struggling to fit in coaching.** Ms. Robichaux and her staff struggle to fit in walkthrough observations because of the large number of teachers at the high school level. They do try to always suggest an area of improvement on their walkthrough observations, and Ms. Robichaux uses the COMPASS rubric as a teaching tool with weaker teachers when she has them observe and score other teachers to see components of the rubric in action.

Mr. Richard provided professional development on the COMPASS rubric and introduced to teachers what the evaluators would be looking for so they "would know what

was expected.” He continues this through informal conversations with teachers by offering suggestions for improvement through email. He always gives struggling teachers comments about how to make their lessons proficient, although he does not conduct formal face-to-face post conferences with his teachers because he feels like communicating electronically is a more efficient way to coach his teachers and improve their practice.

All of the participants mentioned their role as *Instructional Coach* indirectly, and Mr. Boudreaux and Ms. Robichaux directly named themselves as *Instructional Coaches* or *Leaders*. Communicating with their teachers in a variety of ways is important to these principals, and they recognize that one of their responsibilities as a principal is to instruct teachers on their practices. They all do this in using the COMPASS teacher evaluation system in fulfilling their role as *Instructional Coach*.

### **High Stakes Evaluator**

The participants focus on being an *Instructional Coach*, but they also recognized that there is more to being a COMPASS evaluator. They also had to serve in the role of *High Stakes Evaluator* and the pressures that come with that *Position Power*. *Position Power* is how much role-based authority an individual possesses at any given point in time (Fiedler, 1973). The principals in this study had a large amount of *Position Power* because the teacher evaluations they were conducting were directly tied to teacher retention, tenure, and pay. Through their stories, they all exhibited some degree of self-doubt in their role as *High Stakes Evaluator*, and they also admitted to collaborating with other administrators both to increase inter-rater reliability and to build their own confidence as evaluators. They discussed their desire to be fair and accurate evaluators as well as their frustration with the

system, the rubric, and the problems COMPASS has caused in relation to terminating ineffective teachers.

**Experiencing self-doubt.** The participants experienced varying levels of self-doubt in their initial use of the COMPASS evaluation system. Five of the seven participants agreed that their training was not adequate to fully understand and use the rubric to evaluate teachers, and they expressed discomfort in using the rubric initially. Only Mr. Richard felt that he had an appropriate amount of training both within the district and in other districts, but even “with all that training, it’s still not so easy to decide where’s the borderline between highly effective and effective and effective emerging and ineffective.” Mr. Thibodeaux did not want to discuss his training or his feelings about it. The other participants, however, all felt like the training was too short and did not discuss exactly how to score with the rubric in depth. Mr. Breaux feels that schools throughout the LSD are using the rubric differently because of the inadequate training. In his narrative above, he discusses how different principals consider the “Setting Instructional Outcomes” component differently. He also expresses self-doubt in how he is viewing the component:

We have defined it, based on outcomes more so than objectives, but who’s to say that’s even right? You know? I think I’m right...but I definitely know there’s some schools that don’t see it that way.

All of the participants said conducting the first several observations was extremely time-consuming, and those who agreed there was a lack of training blame their insecurities in rating teachers on not enough exposure to in-depth discussion of the rubric. Ms. Fusilier turned to her training manual many times in scoring teachers, “trying to interpret what they meant, taking the book and rereading just to make sure that I was being fair.” Many of

the principals mentioned their desire for fairness and accuracy in using the rubric, yet those who felt unprepared to use the rubric were left frustrated. Ms. Robichaux's narrative also reveals a time when she had to return to her training manual to decide about whether or not scores could be rounded up as her administrative team disagreed about a teacher's rating. She felt frustrated by having one answer from the LSD and another from her training manual. Only Mr. Richard named his primary role as a high stakes evaluator over instructional coach when he self-identified his role in COMPASS as fulfilling the mandated paperwork and deadlines.

In general, the principals seemed much more comfortable and confident in their roles as instructional coaches and less confident as a high stakes evaluator. Although principals in Louisiana have been asked to evaluate teachers for many years now, it is only since the implementation of COMPASS that teacher evaluation has become high stakes and is tied to teacher pay and tenure. With that comes the pressure of assessing teachers' merit based on their performance in the classroom. Although it seems the principals have a general grasp of the policy of COMPASS, most felt uncomfortable in applying the rubric in an actual evaluation setting and blamed this on a lack of training. Donaldson's (2009) study of principals conducting teacher evaluations discovered a lack of training of evaluators and argued that administrators without enough training would not be able to accurately assess teacher effectiveness. The participants in this study echoed this idea as they expressed self-doubt in their shifting role from evaluator to high-stakes evaluator. Feeling uncomfortable with the tool that they are required to use caused the participants to feel somewhat uneasy in this role. The principals used their training and resources to navigate their role as high stakes evaluator, many times leaning on other evaluators at their school site.



**Working together.** All of the participants mentioned working with their fellow administrators at their school site to first use COMPASS. Only Ms. Thibodeaux mentioned working with another principal from a different school within the district to try “to figure out what some of the rubric was really all about.” The other principals only discussed collaboration within their schools in conducting the first several evaluations together and closely reading the rubric together to define what the rubric would look like in practice in their schools. Doing this took a lot of time initially, but several of the participants felt more comfortable in time. Mr. Thibodeaux strives to create a united front with his administrative team at his high school, so they looked at scripting notes to decide how they would each rate lessons. He wants the expectations of teachers to be clear and for their evaluators to be consistent and fair, a sentiment expressed by several participants. The practice of collaboration with other administrators at their school obviously helped to increase their confidence in understanding and using the rubric and developed stronger inter-rater reliability. In-school collaboration is easier to manage in a principal’s busy schedule, rather than trying to fit in time during or after school hours for conversations with other principals. Some principals may also feel insecure about turning to other principals outside of their school because it could cause them to seem incompetent in their role as evaluator among their peers. Schools with only evaluator may turn to out of school collaboration more often or possibly leave the policy to their own interpretations.

The reliance on working with other evaluators at their school cite demonstrates that to the participants, taking on the role of high stakes evaluator felt more comfortable in a group setting with others who were acting in the same capacity. Several principals mentioned a concern for inter-rater reliability in their desire to implement COMPASS fairly.

This same concern with inter-rater reliability is present across several studies of teacher evaluation? (Milanowski and Heneman, 2001; Donaldson and Papay, 2012; Canelake, 2012). Danielson (2011) notes that “most observers require multiple opportunities to practice...to calibrate their judgments with others” (p. 38). Several participants mentioned practicing and discussing scoring within their administrative teams to help build confidence as high stakes evaluators and increase inter-rate reliability as COMPASS evaluators. The participants eased their discomfort with the system by leaning on each other’s experience and expertise to navigate the increased pressure in their new role of high-stakes evaluator.

**Addressing ineffective teachers.** All of the participants also mentioned how it is “impossible to fail” a COMPASS evaluation. For Mr. Boudreaux, even though the worst lesson he has observed was horrible and contained no feedback to students and no student engagement, the teacher still “squeaked out an effective rating.” Mr. Breaux’s worst lesson was also a passing lesson. He recalled:

As bad as that observation was, she scored a 1.8. That’s passing. I don’t know what you have to do to fail this thing. I think you might actually have to fall asleep in the middle of the lesson to fail COMPASS.

Ms. Simoneaux echoed the same feelings: “You have to almost be not breathing to be ineffective.” Ms. Fusilier agrees that it “just takes a lack of so much to be ineffective.” Ms. Robichaux was the most frustrated with COMPASS’s failure to rid schools of ineffective teachers. The state touted her school “as one of the schools that had very few highly effective teachers” meaning that they were using the conservatively, according to the Louisiana Department of Education. She claims “you have to be an idiot to fail it” yet she

cannot “get rid of anybody right now”, not even a teacher who failed her observations and who she “wrote up three times this year” based on policy violations. Ms. Robichaux blames this on the litigation involving COMPASS.

In their roles as high-stakes evaluators, it is evident that the participants felt like their hands were tied when it came to terminating ineffective teachers. Their understanding of the policy led them to believe this is a problem with the COMPASS rubric itself in the cases when teachers who are actually ineffective in the classroom still score as effective and when those scoring ineffective cannot be terminated due to policy litigation. Previous literature reveals an issue in teacher evaluation to be that principals struggled with identifying and scoring the lowest performing teachers due to a lack of training (Donaldson, 2009; Canelake, 2012). This new data evokes the question of whether the underlying issue could perhaps be with the rubrics used for teacher evaluation more than evaluator competency. The participants in Canelake’s (2012) study, just as the participants here, discussed frustration with difficulty to align their rubric with certain instructional practices. Part of the difficulty of the participants fulfilling their role as high-stakes evaluator is their frustration with the rubric and the lack of *Position Power* they feel in this role. In theory, a high-stakes evaluator, as discussed earlier in this section, would yield a high amount of *Position Power* as one who evaluates teachers to determine retention, pay, and tenure. A high-stakes evaluator should be able to identify and terminate an ineffective teacher, but the rubric and current policy leaves the participants feeling powerless when it comes to ineffective teachers. As demonstrated in this narrative, the principals struggle to evaluate teachers whom they feel to be ineffective because they still pass the COMPASS rubric because of its design.

### Revisiting my Subjective I's

Throughout the analysis process, I worked to subdue my own subjective I's that emerged as I composed each participant's story. Several times, my *teacher I* was troubled by some responses from participants as I placed myself in the shoes of the teachers. Mr. Richard, for example, discussed not conducting face-to-face post conferences and not letting teachers know when he walked in that it was their formal COMPASS observation. I found these things unfair; I believe every teacher deserves quality feedback about how to improve their instruction. I found myself feeling badly for those teachers who weren't receiving the coaching that I think should accompany teacher evaluation. I tried to shift my focus to my *evaluator I* who felt the pressures of evaluating teachers within a certain time constraint, and I considered that Mr. Richard was incredibly short-staffed in administration at a large school. This helped me to tell Mr. Richard's story in his own voice without allowing my own emotions to be exposed too much in my interpretations.

Another unexpected finding was that each participant was conducting and utilizing walkthrough data differently at each school. My *teacher I* and my *evaluator I* found this troubling as I personally value inter-rater reliability in teacher evaluation, and I know that this practice alone can so greatly skew evaluation results. When Mr. Breaux and Ms. Robichaux discussed even further inconsistencies they had uncovered with the system, I was outraged that these were not clearly defined and communicated directives by the district. Again, I had to take a step back from my own emotions to convey their stories, highlighting their feelings. In this case, my own feelings helped in the interpretation because they were in agreement with Mr. Breaux and Ms. Robichaux's frustrations with the system.

I allowed my feelings to surface as I collected, interpreted, and analyzed the data during this study because I have been as transparent as possible about my own experiences as an educator. Throughout the process I tried to use subdue my subjective I's when I felt necessary and allowed them to be exposed in other cases when I wanted to inject feelings of my own that were similar to the participants.

### **Summary**

Through their stories, the participants all revealed that they engaged in two main roles through their experiences with COMPASS: *Instructional Coach* and *High Stakes Evaluator*. They communicated both positive and negative feedback to teachers, engaging in both Task Structure and Leader Member Relations. They also felt the pressure of being a High Stakes Evaluator who possessed a large amount of Position Power and experienced the feeling of powerlessness as their hands were tied by current policy. They exhibited self-doubt, frustration, and a desire to be fair and accurate. The principals used both of these roles interchangeably in navigating their responsibilities in Managing the Instructional Program, a dimension of Instructional Leadership Theory.

In Chapter 5, I will revisit the main research question and the original theoretical framework and reflect on its use in the analysis of this data as it relates to the principals' descriptions of their roles in COMPASS teacher evaluation. I will also describe the delimitations of the study as well as implications for theory and practice. I will give recommendations for policy and future research based on the findings that were discovered through my research process.

## **Chapter 5: Discussion**

This final chapter provides an analysis of the findings of how principals describe their role as high-stakes evaluator through their stories about the implementation of COMPASS. It is divided into two sections. The first section revisits the roles that each participant described playing as they implement COMPASS and further analyzes each role and its connection to previous literature. The second section gives implications for theory, practice, policy, and future research on the role of building-level administrators in teacher evaluation.

### **Revisiting the Roles**

#### **Instructional Coach**

The role of instructional coach is one that participants mentioned directly or indirectly, although how they described their specific behaviors within this role varied. All of the participants conducted walkthrough observations and most conducted post conferences, but it surfaced through their stories that each was implementing these practices using various methods. These two practices, however, were a common thread that they all identified as a part of coaching teachers to perform better on COMPASS evaluations. The way in which each participant observed these practices deserves a closer look because it answers the research question of how they define their dual role as both an instructional coach and as a COMPASS evaluator.

**Implementing walkthrough observations.** Classroom walkthroughs are a common practice in schools across the country; however, different models of what constitutes a classroom walkthrough exist (Downey, Steffy, English, Frase, & Poston, 2004; Graf & Werlinich, 2004; Institute for Learning, 2005). The Downey Walkthrough, developed

by principal Carolyn Downey, recognized as the first walkthrough model, consists of several basic components: informal, brief visits of defined reflective areas to gather information about curriculum and instruction with follow-up conversations (Downey, Steffy, English, Frase, & Poston, 2004). On a LearningWalk, taught at the Institute for Learning, administrators spend 10-15 minutes in a classroom analyzing student work, viewing classroom displays, and engaging students in dialogue about what they are learning (Institute for Learning, 2005). Graf and Werlinich's (2004) walkthrough model uses teachers conducting walkthroughs with principals to collect data and provide focused feedback based on standards to one another in a non-judgmental or punitive way. All of these walkthrough models stress that walkthroughs should be non-evaluative or avoid punitive feedback, focusing instead on follow-up professional development that focuses on best instructional practices.

In LSD, it is commonly understood that classroom walkthroughs are unannounced, frequent (approximately once a week, depending on the school), brief (no longer than fifteen minutes) visits to a certain classroom. The protocol for feedback and follow-up varies according to the school and administrator, as evidenced in the narratives of this study. They are generally viewed as an informal data collection method, although teachers do expect timely feedback from the visits. Their informal nature is undoubtedly shifting in the opinion of teachers as now walkthrough data is being used at some schools in combination with COMPASS observation data.

Ovando and Ramirez (2007) identify the monitoring of instruction through walkthrough observations prior to a formal evaluation as a primary function of one who is an instructional leader. This "dry-run" gives the leader an opportunity to work with the

teacher to improve performance prior to high-stakes observation. However, in their study walkthrough data was often not used in combination with a teacher's formal evaluation rating; the data was used to coach the teacher to improve his or her practice before the formal evaluation. The story of the COMPASS evaluators here is much different. As evidenced in Table 2 of Chapter 4, all of the participants conduct walkthrough observations, and six of the seven participants use the data from walkthrough observations in an effort to increase the accuracy of formal observation scores. They valued the fact that walkthroughs increase the sample of teaching behaviors, thereby increasing the reliability of their observation ratings. These same six participants use either all or part of the COMPASS rubric to conduct walkthrough observations. Mr. Boudreaux described the use of walkthrough data in combination with formal evaluation scores as a directive from the district, but it is obvious that either other principals did not receive or interpret the same communication he did or they feel flexibility within the directive to use walkthroughs as they choose for their school. Several participants mentioned that using walkthrough data for a formal observation increases the reliability of their formal evaluation scores because it becomes a more accurate depiction of what is actually happening in the classrooms on a daily basis.

Mr. Boudreaux was the only participant to discuss informal walkthrough observations that are only used as coaching opportunities and not directly tied to COMPASS evaluations. He used his iPad to signal to teachers that he was there collecting data to use for COMPASS and a shake of his head to indicate that he was only there to assess the room as a coach. He was concerned about COMPASS walkthroughs being punitive and not allowing the trust that he wants to have with teachers. It is true that if a



teacher knows the only time a principal walks in his or her room is to evaluate in a way that could affect a score tied to their retention, tenure, and pay, they can begin to view the principal as just an evaluator and not a coach (Donaldson & Donaldson, 2012). Mr. Boudreaux combated this through his informal walkthroughs in which he would give a teacher verbal coaching later that was unrelated to COMPASS scores. To most participants, walkthrough data was an essential part of the evaluation process and helped to inform their coaching. It also was used by some participants, as directed by the district, in combination with formal evaluation scores. The question of whether walkthrough data can function within both purposes at the same time is an idea explored by Mr. Boudreaux when he conducted walkthrough observations with and without his iPad. Evaluators need to clearly understand the purpose of their walkthrough observations in order to use them effectively to coach and/or evaluate teachers.

**Navigating trust.** Several participants, including Mr. Boudreaux, mentioned the importance of trust between himself and teachers. Previous literature tells us that teachers must trust in a principal's training and knowledge to accept coaching and evaluation scores (Milanowski & Heneman, 2001; Donaldson & Donaldson, 2012). This can obviously be a difficult task when principals are asked to observe a wide range of grades and subjects, many outside of the area in which they once taught. Mr. Richard mentioned his efforts to learn other curriculums by studying standards outside his realm of expertise before he conducted formal observations. In his case, because his school had no other administrators, he had to observe the entire faculty. This could become a daunting task for anyone, but in addition to this, he studies the standards and curriculum to feel comfortable to observe his teachers. No other principals discussed a discomfort in knowledge of content areas, but the

participants from the high schools observed teachers by department areas, staying within the departments in which they had teaching experience. This demonstrates that the high school administrative teams considered the difficulty of observing and scoring content area lessons outside of their areas of expertise and assigned evaluators based on this. Again, this can be seen as a move to increase the trust of faculty in the instructional expertise of their evaluator. In the smaller middle school settings, principals were unable to select evaluators with this prior experience. Mr. Thibodeaux and Ms. Fusilier evaluated the same departments two years in a row and were able to coach them from their previous experience in those content areas. They both discussed seeing growth in the teachers they coached and observed. Much research exists in teachers doubting a principals' content and pedagogical knowledge during evaluations (Milanowski & Heneman, 2001; Donaldson & Donaldson, 2012; Lefgren, 2008). Because of the focus on content knowledge and expertise in secondary schools, some of the participants in this study considered this when planning evaluations. The future of teacher evaluation may see increased attention brought on the evaluator's background and own content expertise since this can be a matter of contention when questioning evaluator competency.

**Conducting post-conferences.** All but one of the principals conducted face-to-face post conferences with their teachers following their formal COMPASS evaluations. Several participants considered the post conference to be an important coaching tool. The requirements of COMPASS, as of the 2013-2014 school year, states that the formal observation, or announced observation, must include a pre and post conference and evaluators must provide both areas of commendation and areas of improvement. For the informal, or unannounced observation, the pre and post conference is not required, but

teachers must be provided with the both areas of commendation and areas of improvement (*Louisiana Believes, 2014*). Mr. Richard did not conduct post conferences with his faculty, instead communicating his feedback through email in order to increase efficiency. Mr. Richard is the principal of the largest middle school in the LSD, and as mentioned previously, evaluated the entire faculty himself because his administrative team was missing an administrator during the second year of COMPASS implementation. He stated his primary role in COMPASS to be fulfilling the required documentation and meeting deadlines. He does not identify as a coach, quite possibly because he lacks the resources, i.e. administrative manpower, to spend the time necessary in coaching his teachers.

Mr. Richard was not the only administrator to mention time as an important factor in conducting post conferences. Ms. Simoneaux saw post-conferences as the most important part of her instructional coaching. She spent large amounts of time on each post-conference detailing specific changes teachers could make to their practice. Her previous experience as a TAP master teacher led her to following the TAP post conference protocol, which is aligned with the LDOE guidelines for COMPASS post-conferences because both give areas of strength and weakness. COMPASS does not give a scripted post conference like TAP provides, and the autonomy of post conferences is left to individual school districts. LSD does not mandate a certain protocol for COMPASS post conferences, although a quick Google search pulls up several districts across the state that do provide post conference guidance or scripts. Ms. Simoneaux has seen changes occur in her teachers' practices that she credits to the extensive post conferences in which she coaches her teachers. She also complained about time saying that the school will not run itself and it can be difficult to juggle other administrative responsibilities along with coaching. She names

her previous role as a TAP master teacher to be a time when she had more time to focus on coaching teachers. None of the schools used in this study have master teachers or full-time instructional coaches. Wise et. al (1985) criticized the first teacher evaluations reporting that principals do not have the time necessary to commit to powerful observations and conferences and they should be relieved of some of their managerial tasks in order to commit more time to coaching teachers. A call for master teachers in each school to strengthen teacher evaluation through individualized coaching was also recommended (Wise et. al, 1985). Because time is obviously an issue for the participants in this study as they navigate teacher evaluation in balance with their other responsibilities, some type of master teacher or instructional coach working with principals to coach and support teachers would be beneficial in this district.

**Conclusions on the role of principal as coach.** The participants of this study viewed walkthrough observations and post-conferences as the primary practices within their role as instructional coach. For five out of the seven participants, walkthrough observations were only conducted as data collection tied to COMPASS evaluations. Although tying walkthrough observations to evaluation scores can improve accuracy in scoring, teachers can also view principals only as evaluators and not coaches when this method is used. Observations and evaluations conducted outside of one's area of expertise can also be viewed with mistrust by teachers if they do not believe principals possess the content and pedagogical knowledge necessary to coach or score. Time also becomes a factor in coaching, and one principal did not conduct post conferences and another discussed the difficulty in time management when so much time is spent in post

conferences. Previous literature and this study point to the time consuming nature of evaluation and coaching (Milanowski & Heneman, 2001; Wise et. al, 1985).

### **High-Stakes Evaluator**

The participants of this study were either in their first or second year of COMPASS implementation. All of the participants had been exposed to at least one other teacher evaluation method other than COMPASS, so they were all able to recall previous experiences in their shift from teacher evaluator to high-stakes teacher evaluator. They all expressed some amount of discomfort and self-doubt when it came to using the rubric to score actual lessons. They also all mentioned working with other administrators to increase inter-rater reliability and create a fair evaluation system for their teachers. A more in-depth analysis of these common practices is warranted.

**Using the COMPASS rubric.** All but two of the participants discussed being disappointed in their initial training in implementing the COMPASS rubric. They complained not enough time was spent on each indicator and it left them with questions on how to actually score a real lesson. Previous literature identifies weaknesses in other teacher evaluation systems stem from a lack of adequate evaluator training (Donaldson, 2009; Milanowski & Heneman, 2001; Lefgren, 2008). After her first COMPASS training, Ms. Fusilier was left questioning whether she should have even entered the administrative ranks because she felt so unprepared to use the COMPASS rubric for evaluations. The participants who were dissatisfied with their training did express that after using the rubric for several evaluations they felt more comfortable in using it to score a lesson.

As discussed in Chapter 4, frustrations with using a rubric that may not fit all teaching contexts is not new to teacher evaluation literature. Canelake (2012) also

explored principal's perceptions of teacher evaluation. Unlike this study, however, he found that principals became increasingly frustrated with the rubric and its inability to align with certain instructional practices. The participants of this study found that using the rubric became easier in time as they learned to navigate the new rubric and support teachers in how to achieve the expectations of the rubric. This trend during the first two years of COMPASS implementation could indicate that over time, principals will become more comfortable and confident in using COMPASS as an evaluative tool and teachers will increase in effectiveness according to the rubric.

What did not improve in time for most of the participants were the holes or gaps they described in the rubric. Mr. Boudreaux is frustrated with the actual COMPASS rubric and its lack of "alignment." He finds that each component contains attributes that "are disjointed and not related to one another attributes." To him, this is because the COMPASS rubric is not the entirety of Danielson's Framework, and therefore is missing pieces. He began using COMPASS during the pilot in another district when it did consist of the entirety of the rubric, so although other principals felt that there were holes or missing pieces, Mr. Boudreaux is the only participant who could actually compare the five-part COMPASS rubric to the entire Framework. Ms. Simoneaux also found it "difficult to match up the evidence" and mentioned, as did four other participants, that a lack of adequate training also led to their frustration in using the rubric. Ms. Robichaux finds that "there are some things, they may have it under the emerging level, but then you don't see it on the proficient level." Mr. Boudreaux claims this is because the COMPASS rubric is not a rubric at all and the indicators, in his opinion, were never intended to be used to evaluate teachers. He feels the Framework was originally intended to be a guide for teachers to

implement as best practices, not a set criteria for evaluation. These responses to using the rubric in the voices of the participants answer the problem statement for this study: Charlotte Danielson helped Louisiana create the COMPASS rubric yet warned that not using “the full instrument...decreases accuracy.” The participants of this study obviously feel the discrepancy of not using the entire Framework to evaluate teachers. During Louisiana’s pilot for COMPASS, the piloting districts used the entire Framework, but the LDOE decided the process became too time-consuming for principals. Mr. Boudreaux, who participated in the pilot, described it as cumbersome but then criticized the shortened COMPASS rubric as not enough to accurately assess a lesson. Charlotte Danielson was also concerned that Louisiana did not pilot what is now the COMPASS rubric after the initial pilot using the entire Framework. This raises the question of what the COMPASS rubric might look like today if it was piloted? Could there have been a happy medium between using the full Framework and what is now the COMPASS rubric? A rubric that principals don’t find overwhelming or daunting, yet one they feel is comprehensive enough to evaluate teachers without the gaps or holes described by the participants in this study? No revisions have been made to the rubric at this point, although the Louisiana Department of Education continues to release documents to help principals in using the rubric, such as criteria they can look for in English Language Arts and Math that fit the Common Core shifts in classroom instruction.

Prior to COMPASS, the vast majority of teachers in Louisiana passed their evaluations with an effective rating, a major selling point of the COMPASS legislation, Ironically, this has not changed with COMPASS. Even though the COMPASS system has tiered teacher’s effectiveness, the majority of teachers still pass their evaluations.

According to the participants in this study. Another reoccurring issue with the rubric throughout interviews was the difficulty COMPASS presented in the termination of ineffective teachers. This happened either because the rubric made it difficult to classify a teacher as ineffective and allowed them to scrape by with a low effective rating or those who were deemed ineffective on the rubric could not be terminated because of the lawsuits already involving termination of teachers based on teacher evaluation scores. As discussed in Chapter 4, this led to the participants feeling powerless in their roles as high stakes evaluators. It also leads to a question about the COMPASS rubric itself: Why is it so difficult to fail a COMPASS observation, as the participants in this study report? The indicators in the ineffective category do basically call for a lesson to be completely abysmal in each of the five areas. The participant's complaint is that a teacher is ineffective even if they are terrible in only some of the five areas, yet if they pass any of the areas, it can give them an overall passing score. This again begs the question of piloting the newer, shortened rubric. Surely, if it is a complaint in the small sample size of seven in this study, it would have been an issue heard from many principals across the state and the ineffective category could have been reworked to give a more accurate score for those teachers.

**An unintended learning community.** Neither the LDOE nor the LSD mandated any way to ensure or increase inter-rater reliability or have principals work together in learning the rubric, yet all of the participants in this study discussed working with other administrators to learn how to implement COMPASS. Their self-doubt in implementing the rubric caused them to naturally be drawn to working with others in the same circumstances as themselves. The participants also cared about implementing COMPASS in a way that was fair for their teachers, so they wanted teachers to receive similar scores no



matter their evaluator. Previous literature discusses inter-rater reliability and inconsistency among evaluators as an issue with other teacher evaluation systems (Milanowski & Heneman, 2001; Donaldson and Papay, 2012; Canelake, 2012). The participants in this study discussed working together to learn the rubric and feel more comfortable with using the rubric in evaluations. This unintended yet positive consequence of COMPASS forced principals to work together as they learned. Beabout (2012) calls the collaborative process after an organization experiences turbulence (i.e. initial implementation of COMPASS) perturbation. He argues for the importance of perturbation in the survival of learning organizations as it forces individuals to work together to establish cultures of collaboration. In this study, the principals conducted brief and extended teacher observations and discussed the rubric in-depth. The participants did not discuss if this would be a practice that would be continued in the future since they now feel more comfortable with the rubric. Perhaps they would conduct walkthrough observations and some initial evaluations with new administration as a training exercise, but in order to maintain inter-rater reliability from year to year, this practice would need to be continued (Danielson, 2011). Without this practice as an LDOE or LSD mandate or guideline, there is a possibility it will be discontinued at many schools because of the time constraints the participants are already facing. Fink (2003) explored the unintended consequences of top-down reform. In a secondary school, he found increased pressure, isolation, and time limitations to cause principals to potentially seek alternate career paths. If principals in Louisiana feel the same pressure, isolation, and time constraints, all of which were mentioned by participants in this study, an unintended consequence of COMPASS could be the exodus of some principals from the career.

**Conclusions on the role of principal as high stakes evaluator.** The participants in this study identified themselves as instructional coaches more than high stakes evaluators, but being a high stakes evaluator is an obligatory role and they perform actions as an inherent part of the role. They have to use the COMPASS rubric even though they did not feel adequately trained and they have issues with the comprehensiveness of the rubric in evaluating a lesson. They have naturally chosen to work together to subdue their self-doubts in using the rubric, either intentionally or subconsciously increasing the inter-rater reliability at their schools. We are left with questions within these roles like: How would their actions be different if the current COMPASS rubric would have been piloted? Will their focus on working together change as they become more comfortable with the rubric? How will future policy impact and change their roles in high stakes teacher evaluation?

### **Implications for Theory**

The original theoretical framework for this study proposed that additions needed to be made to Hallinger and Murphy's (1985) Instructional Leadership Model. It was proposed that Hersey and Blanchard's (1993) Contingency Theory Model could help to fill a gap found in the Instructional Leadership Model. The participants in this study did engage in both task and relationship behaviors as they implemented COMPASS. They demonstrated task behaviors as they coached the teachers on their performance on walkthrough observations and through post conferences. Several participants noted that trust was a very important factor because their teachers needed to trust their pedagogical skills to coach them, thus demonstrating relationship behavior.

In Hallinger and Murphy (1985) model, supervising evaluation and instruction is a descriptor that falls into the dimension of Managing the Instructional Program. Although

the model includes this mention of evaluating teachers, it does not mention coaching teachers' practices. This study demonstrates that instructional coaching is directly tied to teacher evaluation. The role of instructional coaching encompasses those practices that were both task and relationship behaviors. Most of the participants named instructional coaching as their primary role in teacher evaluation, and Mr. Thibodeaux went so far as to describe his teachers' scores as being a direct representation of his own leadership abilities. Adding a descriptor to Managing the Instructional Program about instructional coaching would update the existing model in a way that is actually applied by principals, who identify in the roles of both high stakes evaluator and instructional coaches, implementing high-stakes teacher evaluations.

Fred Fiedler's (1973) Contingency theory also informed data analysis in this study. *Position Power* is one of the three parts of this theory and refers to the amount of power that the leader yields over group members over a specific task. In this case, how much power the participants have over the teachers based on how they perform on a COMPASS evaluation. I used the idea of Position Power to analyze the findings about principals as they take on the role of high stakes evaluator. As discussed in Chapter 4, one would assume that a principal acting in the role of high stakes evaluator would possess a large amount of Position Power because the scores from his or her evaluations are used for teacher retention, tenure, and pay. At first glance, it would seem that the participants of this study experienced a significant increase in position power, seeing as COMPASS gives individual principals the power to terminate ineffective teachers and reward highly effective teachers with increased pay. That is, after all, the whole idea behind COMPASS. What this study reveals though, is that the participants actually felt their hands were tied when it came to

scoring teachers as ineffective based on the rubric or actually terminating teachers who were rated ineffective because of policy litigation. This demonstrates that there is a difference between *assumed* Position Power and *perceived* Position Power. Most of the participants' teachers and probably most teachers in the state believe that their principals have a large amount of power in their ability to terminate them, but in reality, the participants in this study felt the opposite was true. They felt confident in their abilities to identify ineffective teachers but because of the dichotomy between the rubric and actual classroom implementation, they were unable to score teachers as ineffective, even when they felt they should be. The one participant who actually did score a teacher as ineffective could not remove the teacher because of the ongoing legal issues already involving COMPASS at the state level.

### **Implications for Practice**

Several practices were demonstrated by one or more of the participants in this study that could impact the accuracy of COMPASS teacher evaluation and simply improve teacher evaluation in general. Conducting informal walkthrough observations outside the realm of data collection for COMPASS can help to increase trust from teachers. Comparing student achievement results to observation results can help to inform practice and future policy. Continuing and expanding the practice of conducting walkthrough and formal observations together can increase inter rater reliability.

Mr. Boudreaux would tell teachers in which role he was acting when he entered a room with or without his iPad to conduct a walkthrough observation. He wanted teachers to trust his coaching and not always feel like they were being judged by an evaluator every time he stepped in the room. Previous literature identifies a lack of trust of one's evaluator

as an issue that causes teacher evaluation to be unsuccessful in actually changing classroom practices (Milanowski & Heneman, 2001; NEA, 2010; Donaldson & Donaldson, 2012). If the goal of COMPASS is to provide teachers with meaningful feedback that they can use in their classrooms, then they must trust that their evaluators are competent coaches. By conducting walkthrough observations that are not scored, principals can build relationships with their teachers that are based on more than just formal evaluation data. Although data from walkthrough observations can be used to strengthen the accuracy of COMPASS evaluation results, it is important for principals to at times be present in classrooms without wearing the hat of high stakes evaluator.

Ms. Fusilier discussed how she examines student achievement results in relation to the COMPASS evaluation results of her teachers. This is an important practice for several reasons. In her case, she realized that she was evaluating a teacher who may not get highly effective evaluation scores because he is not a fancy flashy teacher; however, his student achievement results are solid. This can tell her one of two things: 1. He could do even better and push his students even further if he used some different instructional techniques because he is already a good teacher or 2. He is a highly effective teacher, and his evaluation scores just maybe don't show it because of the rubric. Either way, she can use all of the data from this teacher to help her make decisions about her school. She discussed how his students see a lot of growth in their ACT score during his course, so when she needed a chair for the ACT committee, she knew to ask him. If she would have only relied on her observation scores, she may not have realized that he was well-suited for this role. Ms. Fusilier stated the importance of seeing the big picture of teacher evaluation, and this is a practice that all principals should use. Authentic assessment of teachers begins with using

common sense to evaluate all the information one has on that teacher to drive decision-making. Mr. Boudreaux also mentioned that the data from COMPASS helped him to make decisions on future professional development for his teachers. Because his teachers did not score well in the questioning component of the rubric, he planned professional development to address teacher and student questioning in instruction. Principals should utilize all data available on a teacher's practice, including COMPASS data, in making staffing and professional development decisions.

All of the participants in this study discussed collaborating with other administrators to improve their use of the COMPASS rubric. Only one participant discussed the rubric with an administrator at another school in the LSD. The participants worked together during year one and year two of COMPASS implementation, but as year three approaches, it is still important to conduct some walkthrough and formal observations together to ensure inter-rater reliability (Danielson, 2011). School districts should even encourage administrators within the district to visit other schools and conduct walkthrough observations together. Mr. Breaux was concerned that administrators at other schools throughout LSD were evaluating the Setting Instructional Outcomes part of the rubric differently. Conducting district walkthroughs in which principals visited other schools could increase inter-rater reliability as a district and ensure collaboration between principals strengthens the evaluation process.

### **Implications for Policy**

The findings of this study suggest several possible policy changes for Louisiana teacher evaluation. Two of the major issues that were revealed in this study have already been addressed by the LDOE in changes to COMPASS. Ms. Robichaux did not like that

COMPASS forced her to evaluate every teacher in her school twice a year, no matter their level of effectiveness. As of the 2013-2014 school year, revisions to Bulletin 130 (Regulations for the Evaluation and Assessment of School Personnel) removes the distinction between formal and informal evaluations, allowing greater flexibility in the type and duration of observation be left up to individual school districts. Revisions to Bulletin 130 also allows principals to evaluate Highly Effective teachers less often and observe lower performing teachers more often throughout the school year. Ms. Robichaux complained that she and her administrative team spent so much of their time evaluating everyone that they did not have a chance to work closely with lower performing teachers. Autonomy in the number of observations should help Ms. Robichaux and other principals who struggle with accommodating the demands of a large faculty.

Five of the seven participants did not like the current COMPASS rubric and felt like their training on the rubric was inadequate in actually preparing them for scoring teachers using the rubric. As the problem statement addresses, the pilot for COMPASS included the entire Framework and the current COMPASS rubric only includes five parts of the twenty-two parts. The participants in this study felt like this left gaps in the rubrics that made it difficult to use at times. Policy changes need to be made that address these issues with the rubric and feedback from principals should be used to further align the rubric in a way that is easier to use. This could be done through a pilot in which feedback from several districts is used to adjust the rubric. Further training should also be developed by the LDOE and local school districts to provide clarity on the inconsistencies revealed in this study.

### **Limitations of the Study and Future Research**

Although this study does yield results that inform theory, practice, and policy about teacher evaluation, it does have limitations and it does not address many of the gaps in the existing literature about teacher evaluation. A major limitation of this study is because it is a narrative study, the sample size of seven participants is small, and so the results here perhaps may not be translated to a larger population. It also took place only in one school district in Louisiana, so the results may not be the same that could possibly be found in other school districts across the state. This study also only included middle and high schools in which all teachers were departmentalized within content areas, so results of an elementary school may be very different.

Although many studies have been conducted over the past century on teacher evaluation (Canelake, 2012; Halverson, Kelly, & Kimball, 2004; Milanowski, Kimball, & White, 2004; Kane, Taylor, Tyler, & Wooten, 2010; Schacter & Thum, 2004; Daley & Kim, 2010), a great increase in policies for high stakes teacher evaluations has occurred over the last decade, so there is much more to be explored. This study tried to answer the question of how principals describe their role as high-stakes evaluator through their stories of implementing COMPASS. Two reoccurring questions have appeared in previous literature and this study: rubric reliability and a correlation between student achievement data and teacher evaluation results. The participants in this study questioned the use of the rubric and in previous studies complained about the use of their respective rubrics and its alignment to the overall goal of teacher evaluation (Canelake, 2012; Halverson, Kelly, & Kimball, 2004). Previous literature has shown mixed results of correlations between student achievement data and teacher evaluation results (Milanowski, Kimball, & White,



2004; Kane, Taylor, Tyler, & Wooten, 2010; Schacter & Thum, 2004; Daley & Kim, 2010). Future research could include comparing Louisiana's SLT results with COMPASS results.

Further studies should also include two new questions emerging from this study unaddressed in previous literature. First, walkthrough observations became a focal point of this study as participants mentioned them multiple times throughout data collection. Walkthrough observations in relation to formal evaluations is an unexplored area of teacher evaluation literature. Another idea mentioned in previous literature indirectly is that of principals conducting evaluations outside of their areas of content expertise. It was noted in this study that the high school principals conducted evaluations on teachers within the content areas they used to teach. A study exploring how this impacts coaching and evaluation could potentially reveal results for future practice and policy in teacher evaluation.

## References

- Act 54, (2010). Louisiana State Legislature.
- Ahn, T. & Vigdor, J. (2013). Were all those standardized tests for nothing? The lessons of no child left behind. *American Enterprise Institute*.
- Alicias, E.R. (2005). Toward an objective evaluation of teacher performance : The use of variance partitioning analysis. *Education Policy Analysis Archives, 13*(30), 1-15. Retrieved from ERIC database.
- American Press (2012). Teacher union challenges Jindal education changes. Retrieved from <http://www.americanpress.com/Teachers-union-challenges-Jindal-education-changes>
- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system (EVAAS). *Educational Researcher, 37*(2), 65-75. Doi: 10.3102/0013189X08316420
- Baker, E.L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J., & Shepard, L.A. (2010). Problems with the use of student test scores to evaluate teachers. Washington, D.C.: Economic Policy Institute. Retrieved from <http://epi.org/publications/entry/bp278>
- Barth, R. (1986). On sheep and goats and school reform. *Phi Delta Kappan, 68*(4), 293-296.
- Blase, J., & Blase, J. (1999). Effective instructional leadership: Teachers' perceptions on how principals promote teaching and learning in schools. *Journal of Educational Administration, 38*(2), 130-141.
- Beabout, B. R. (2012). Turbulence, perturbation, and educational change. *Complicity: An International Journal of Complexity and Education, 9*(2), 23-37.
- Bloor, M., and Wood, F. (2006). *Keywords in qualitative methods*. Thousand Oaks, California: Sage.
- Callahan, R.E. (1962). *Education and the cult of efficiency*. Chicago: The University of Chicago Press.
- Canelake, C. (2012). Implementing a standards-based teacher evaluation system: Learning experiences for administrators in an urban school district (Doctoral dissertation). Retrieved from The Carnegie Project on the Education Doctorate.
- Chase, S.E. (2005). Narrative inquiry: Multiple lenses, approaches, voices. In N.K. Denzin & Y.S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (3<sup>rd</sup> ed.) (pp. 651-674). London: Sage.

- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*, third edition. Los Angeles: Sage.
- Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review*, 54(2), 129-151.
- Czarniawska, B. (2004). *Narratives in social science research*. Thousand Oaks, California: Sage.
- Daley, G. & Kim, L. (2010). A teacher evaluation system that works. *National Institute for Excellence in Teaching* working paper.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, Va: Association for Supervision and Curriculum Development.
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- The Danielson Group (2011). The framework for teaching. Retrieved from <http://www.danielsongroup.org/article.aspx?page=frameworkforteaching>
- Darling-Hammond, L., Wise, A.E., & Pease, S.R. (1983). Teacher evaluation in the organization context: A review of the literature. *Review of Educational Research*, 53(3), 285-328.
- Donaldson, M.L. (2009). So long Lake Wobegon?: Using teacher evaluation to raise teacher quality. Retrieved from Center for American Progress website: <http://www.americanprogress.org>
- Donaldson, M.L., & Donaldson, Jr., G.A. (2012). Strengthening teacher evaluation: What district leaders can do. *Educational Leadership*, 69(8), 78-82.
- Donaldson, M.L. & Papay, J.P. (2012). Reforming teacher evaluation: One district's story. *Center for American Progress*.
- Downey, C., Steffy, B., English, F., Frase, L., & Poston, W. (2004) *Three minute classroom walkthrough*. Thousands Oaks, CA: Corwin Press.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37, 15-24.
- Elliot, J. (2005). *Using narrative in social research: Qualitative and quantitative approaches*. Thousand Oaks, California: Sage.

- Etherington, K. (2004). *Becoming a reflexive researcher: Using our selves in research*. London: Jessica Kingsley Publishers.
- Fiedler, F. (1973). The contingency theory and the dynamics of leadership process. *Advances in Experimental Social Psychology*, 11, 60-112.
- Fink, D. (2003). The law of unintended consequences: The 'real' cost of top-down reform. *Journal of Educational Change*, 4(2), 105-128.
- Fuller, E. (2006). Teacher and parent perceptions of the Texas assessment of knowledge and skills. Retrieved from <http://www.atpe.org/advocacy/issues/TAKSStudy.pdf>
- Fuller, E. & Hollingworth, L. (2013). A bridge too far? Challenges in evaluating principal effectiveness. *Educational Administration Quarterly*. DOI: 10.1177/0013161X13506595
- Garland, S. (2012, November 20). Teacher evaluation architect warns of lawsuits against Louisiana's new system. *The Hechinger Report*.
- Glesne, C. (2011). *Becoming qualitative researchers: An introduction*. Boston: Pearson.
- Graf, O., & Werlinich, J. (n.d.). Observationfrustrations. Is there another way? The walkthrough observation tool. Retrieved from <https://qcc.wikispaces.com/file/view/Walkthroughs+-Observation+Frustrations.pdf>
- Hallinger, P., Bickman, L., & Davis, K. (1996). School context, principal leadership and student achievement. *Elementary School Journal*, 96(5), 498-518.
- Hallinger, P., & Murphy, J. (1985). Assessing the instructional leadership behavior of principals. *Elementary School Journal*, 86(2), 217-248.
- Halverson, R., Kelley, C., & Kimball, S. Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. *Educational Administration, Policy, and Reform: Research and Measurement*, 153-188.
- Hersey, P., & Blanchard, K.H. (1993). *Management of organizational behavior: Utilizing human resources* (6<sup>th</sup> ed.). Englewood Cliffs, NJ: Prentice Hall.
- Hoy, W.K., & Miskel, C.G. *Educational administration: Theory, research, and practice* (8<sup>th</sup> ed.). Boston: McGraw Hill.
- Institute for Learning (n.d.) *The LearningWalk<sup>sm</sup>*. Retrieved March 1, 2006, from <http://www.instituteforlearning.org/howwk.html>

- Jacob, B.A. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- Kane, T.J., Taylor, E.S., Tyler, J.H., & Wooten, A.L. (2010). Identifying effective classroom practices using student achievement data. *National Bureau of Economic Research* working paper.
- Kersten, T.A. & Israel, M.S. (2005). Teacher evaluation: Principals' insights and suggestions for improvement. *Planning and Changing*, 36(1-2), 47-67.
- Kersting, N.B., Chen, M., & Stigler, J.W. (2013). Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Education Policy Analysis Archives*, 21(7), 1-39. Retrieved from <http://epaa.asu.edu/ojs/article/view/1167>
- Kimball, S.M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.
- Leithwood, K., & Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research*, 52(3), 309-339.
- Lewis, A.C. (1982). Evaluating educational personnel. Arlington, Va: American Association of School Administrators.
- Louisiana Department of Education (2012). Act 1 Compensation. Retrieved from <http://www.louisianabelieves.com/docs/teaching/act-1-legislative-memo.pdf?sfvrsn=2>
- Marion, R. (2002). *Leadership in education: Organizational theory for the practitioner*. Upper Saddle River, NJ: Prentice Hall.
- McGreal, T.L. (1982). Effective teacher evaluation systems. *Educational Leadership*, 303-305.
- Milanowski, A.T. & Heneman, H.G. III. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education*, 15(3), 193-212.
- Milanowski, A.T., Kimball, S.M., & White, B. (2004). The relationship between standards-based teacher evaluation scores and student achievement: Replication and extension at three sites. *Consortium for Policy Research in Education*.

- National Education Association (2010). Teacher assessment and evaluation: The National Education Association's framework for transforming education systems to support effective teaching and improve student learning. Retrieved from National Education Association website: <http://www.nea.org/home/41858.htm>
- Ovando, M.N., & Ramirez, A. (2007). Principals' instructional leadership within a teacher performance appraisal system: Enhancing students' academic success. *Journal of Personnel Evaluation in Education*, 20, 85-110.
- Parsons, T. (1960). *Structure and process in modern society*. Glencoe, NY: The Free Press.
- Peshkin, A. (1988). In search of subjectivity – One's own. *Educational Researcher*, 17(7), 17-21.
- Popham, W.J. (1973). Alternative teacher assessment strategies. Working paper for a meeting of the Multi-State Consortium on Performance-Based Teacher Evaluation. New Orleans, Louisiana, February 1973.
- Riessman, C.K. (2000). Analysis of personal narratives. In J.F. Gubrium & J.A. Holstein (Eds.), *Handbook of Interviewing*. New York: Sage.
- Rogers, C.B., & Weems, D.M., (2010). Are US teachers making the grade?: A proposed framework for teacher evaluation and professional growth. *Management in Education*, 24(1), 19-24.
- Sartain, L., Stoelinga, & Brown, E.R. (2011). Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Consortium on Chicago School Research as the University of Chicago Urban Education Institute.
- Schacter, J. & Thum, Y.M. (2004). Paying for high- and low- quality teaching. *Economics of Education Review*, 23, 411-430.
- Stumbo, C., & McWalter, P. (2011). Measuring effectiveness: What will it take? *Educational Leadership*, 68(4), 10-15.
- Tennessee Department of Education. (2012). Teacher evaluation in Tennessee: A report on year 1 implementation. Retrieved from [http://www.tn.gov/education/doc/yr\\_1\\_tchr\\_eval\\_rpt.pdf](http://www.tn.gov/education/doc/yr_1_tchr_eval_rpt.pdf)
- U.S. Department of Education (2009). Race to the Top: executive summary. Washington, DC: U.S. Department of Education. Retrieved November 17, 2010.
- Wise, A.E., Darling-Hammon, L., McLaughlin, M.W., & Bernstein, H.T. (1985). Teacher evaluation: A study of effective practices. *The Elementary School Journal*, 86(1), 60-121.

Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67.

## Appendix A

Dear Research Participant:

The dissertation study, “Stories of Principals’ Roles as a Result of High Stakes Teacher Evaluations in Louisiana” involves research on the experience of principals in schools as they conduct COMPASS teacher evaluations. Should you choose to participate, you will be asked to participate in 3 interviews over the next year. Each interview will last approximately 1 hour, making for a total of 3 hours in the next year.

Your participation in this study entails some risk relating to employability and reputation, although the chance of this risk is slight. Because you will be discussing COMPASS teacher evaluations in use at your place of employment, your participation will be kept confidential and your real name will not be used in any publications created from this research. Participation in this study is voluntary and refusal to participate will involve no penalty or loss of benefits to which you are otherwise entitled. You may discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled.

This research could benefit you in that it may cause you to reflect upon your role as a COMPASS teacher evaluator. I intend this study to be used to inform school leaders about teacher evaluation practices and policymakers about teacher evaluation policy in practice. Your insights are essential to providing a better sense of how we can improve the practice of teacher evaluation in a way that benefits school leaders, teachers, and students.

To maximize confidentiality, neither your name nor your school’s name will be used in any the publications resulting from this research. Interview will be audio recorded and will be kept secure and will only be accessible by JennaLynn Chiasson, the researcher. If you have any questions about this particular study, please contact Dr. Brian Beabout at (504) 280-7388 or [bbeabout@uno.edu](mailto:bbeabout@uno.edu). If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact Dr. Ann O’Hanlon at the University of New Orleans at 504-280-3990.

---

Participant (print name)

---

Researcher (print name)

---

Participant (sign)                      date

---

Researcher (sign)                      date



Appendix B  
**Questionnaire**

Job Title: \_\_\_\_\_

Years of experience in education: \_\_\_\_\_

Years in current position (at this school site): \_\_\_\_\_

Years of teaching experience: \_\_\_\_\_

Years of administrative experience: \_\_\_\_\_

Gender: \_\_\_\_\_ Race: \_\_\_\_\_

Age: (please circle one)

20-30

31-40

41-50

51-60

61-70

Number of teachers you are responsible for conducting COMPASS evaluations on annually: \_\_\_\_

Number of COMPASS evaluators at your school site: \_\_\_\_\_

Total number of COMPASS evaluations you have conducted: \_\_\_\_\_

## **Interview Protocol #1**

Tell me about the first COMPASS evaluation that you conducted.

Tell me about the most recent COMPASS evaluation that you conducted.

Tell me about the best COMPASS evaluation that you conducted.

Tell me about the worst COMPASS evaluation that you conducted.

Tell me about the most comfortable COMPASS evaluation that you conducted.

Further prompts if necessary:

What made it the best?

What made it the worst?

Why was it the most comfortable?

## **Interview Protocol #2**

Tell me about any COMPASS evaluations you have conducted since our first interview.

Describe your role in COMPASS teacher evaluation.

Tell me about how you were introduced to COMPASS.

What do you see as the purpose of COMPASS?

What abilities have you used in your role as a COMPASS evaluator?

Describe the interrelationship between your role as principal and your role as evaluator.

### **Interview Protocol #3**

\*Interview 3 protocol will take place after Chapter 4 is drafted. Each participant will be asked to member check their section of the chapter for accuracy in interpretation.

Appendix C



INFORMED CONSENT for RESEARCH

Title of Study: Stories of Principals' Roles as a Result of High Stakes Teacher Evaluations in Louisiana

Dear \_\_\_\_\_:

I am a graduate student under the direction of Professor Brian Beabout in the Department of Educational Administration at the University of New Orleans. I am conducting a research study to fulfill the requirements of a doctoral dissertation. The purpose of this study is reveal stories of secondary principals about their roles in COMPASS teacher evaluation at their school sites.

I am requesting your participation, which will involve three in person interviews, not lasting longer than ninety minutes each. Interviews will be scheduled at your convenience. Your participation in this study is voluntary. If you choose not to participate or to withdraw from the study at any time, there will be no penalty. The results of the research study may be published, but your name will not be used.

Participating in this study allows you to contribute to the field of research in teacher evaluation. Principal voices in this field of research are vitally important, making your participation in this study extremely valued.

If you have any questions concerning the research study, please call me at (985) 413-0324 or Dr. Brian Beabout at (504) 280-7388.

Sincerely,

JennaLynn Galjour Chiasson

By signing below you are giving consent to participate in the above study.

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Printed Name

\_\_\_\_\_  
Date

If you have any questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, please contact Dr. Ann O'Hanlon at the University of New Orleans (504) 280-6501.

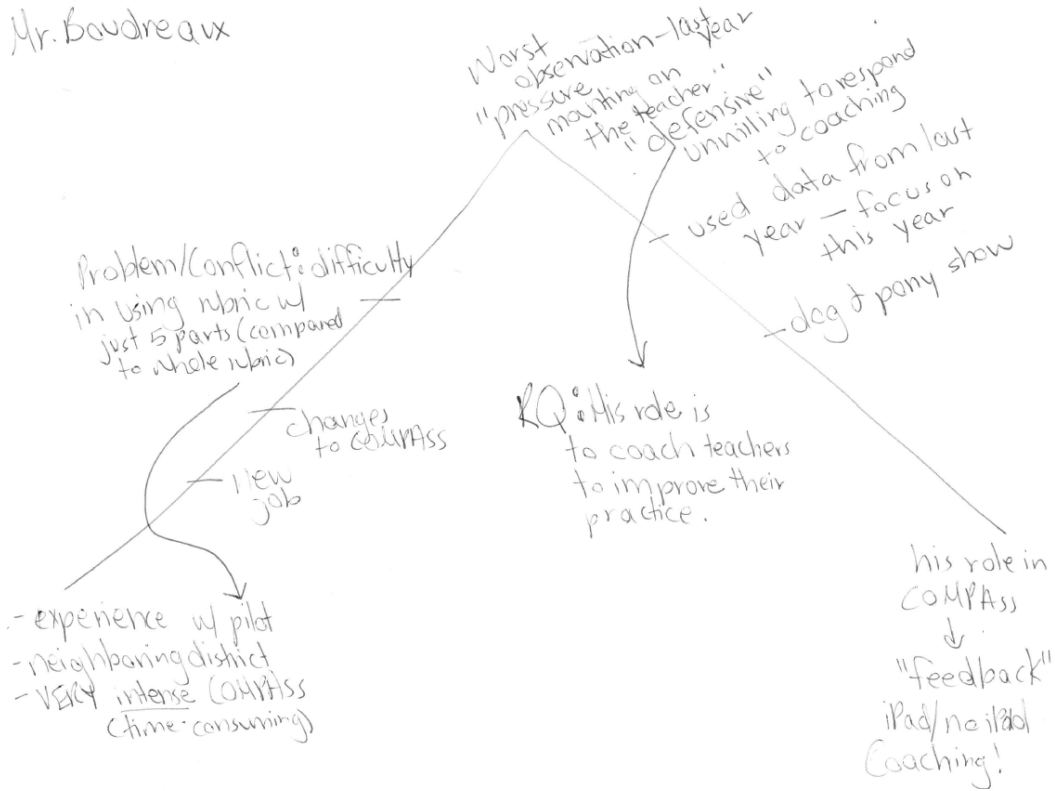
## Appendix D

*How do principals describe their roles as high-stakes evaluator through the implementation of COMPASS?*

<b>Participant</b>	<b>Answer to RQ</b>	<b>Hallmarks of story</b>	<b>Current order</b>
Mr. Boudreaux	His role is to coach teachers to improve their practice.	iPad vs. no iPad (walkthroughs), coaching teachers even more than required	1
Mr. Breaux	His role is to implement teacher evaluation as fairly and accurately as possible.	Former coach, wanting to follow the rules	2
Ms. Robichaux	Her role is forced. She must perform it despite her bitterness over the vagueness of implementation.	Liked the “old rubric” used it to get rid of teachers	3
Ms. Richard	Her role is to fulfill what she is mandated to do by COMPASS, nothing more.	She does not script, emails scores with no post conferences	4
Ms. Simoneaux	Her role is to impact teaching practice through her feedback.	TAP trained, strong emphasis on post conferencing	5
Mr. Thibodeaux	His role is to support teachers. He evaluates his own leadership through COMPASS.	Former PE teacher/coach who is still learning about classroom instruction	6
Ms. Fusilier	Her role is to evaluate teachers in a way that is reflective and considerate of all decisions.	Uses VAM data to “confirm her thinking” – not always accurate	7

Appendix E

Mr. Boudreau



## Appendix F

### Coding Scheme:

IF – Impossible to Fail

SD – Self-Doubt

Fr – Frustration

WT – walkthrough

WTog – Working together

T – Time

F – Fairness

TB – Task Behavior

RB – Relationship Behavior

TD – Teacher Defensiveness

TA – Teacher Acceptance

C – Teacher Coaching

Exp – Exposition, a story's beginning, i.e. characters, setting

Conf – Conflict

RA – Rising Action

Cl – Climax

FA – Falling Action

Conc – Conclusion



## Appendix G

**University Committee for the Protection  
of Human Subjects in Research  
University of New Orleans**

---

*Campus Correspondence*

Principal Investigator: Brian Beabout  
Co-Investigator: JennaLynn Chiasson  
Date: March 27, 2014  
Protocol Title: "Stories of Principals' Roles as a Result of High Stakes  
Teacher Evaluations in Louisiana"  
IRB#: 02Mar14

The IRB has deemed that the research and procedures are compliant with the University of New Orleans and federal guidelines. The above referenced human subjects protocol has been reviewed and approved using expedited procedures (under 45 CFR 46.116(a) category (7)).

Approval is only valid for one year from the approval date. Any changes to the procedures or protocols must be reviewed and approved by the IRB prior to implementation. Use the IRB number listed on this letter in all future correspondence regarding this proposal.

If an adverse, unforeseen event occurs (e.g., physical, social, or emotional harm), you are required to inform the IRB as soon as possible after the event.

Best wishes on your project!

Sincerely,



Robert D. Laird, Ph.D., Chair  
UNO Committee for the Protection of Human Subjects in Research

## VITA

The author was born in Raceland, Louisiana. She obtained her Bachelor's degree in English Education from Nicholls State University in 2009. She completed her Master of Education degree as a Reading Specialist from Nicholls State University in 2011. She joined the University of New Orleans education graduate program to pursue a Ph.D. in Educational Administration in 2011.